Interdisciplinary Applied Mathematics 19

Marek Kimmel
David E. Axelrod

Branching Processes in Biology

Second Edition

Springer

Interdisciplinary Applied Mathematics

Problems in engineering, computational science, and the physical and biological sciences are using increasingly sophisticated mathematical techniques. Thus, the bridge between the mathematical sciences and other disciplines is heavily traveled. The correspondingly increased dialog between the disciplines has led to the establishment of the series: Interdisciplinary Applied Mathematics.

The purpose of this series is to meet the current and future needs for the interaction between various science and technology areas on the one hand and mathematics on the other. This is done, firstly, by encouraging the ways that mathematics may be applied in traditional areas, as well as point towards new and innovative areas of applications; and secondly, by encouraging other scientific disciplines to engage in a dialog with mathematicians outlining their problems to both access new methods as well as to suggest innovative developments within mathematics itself.

The series will consist of monographs and high-level texts from researchers working on the interplay between mathematics and other fields of science and technology.

More information about this series at http://www.springer.com/series/1390

Marek Kimmel • David E. Axelrod

Branching Processes in Biology

Volume 19

Springer

Marek Kimmel
Department of Statistics
Rice University
Houston
Texas
USA

David E. Axelrod
Department of Genetics
Rutgers University
Piscataway
New Jersey
USA

Mathematics Subject Classification: 91B70, 92CXX, 92C37, 92DXX, 92D10, 92D15, 92D20, 92D25, 93C95, 60J80, 60J85

ISSN 0939-6047 ISSN 2196-9973 (electronic)
ISBN 978-1-4939-1558-3 ISBN 978-1-4939-1559-0 (eBook)
DOI 10.1007/978-1-4939-1559-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014948195

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Barbara and Helena

Preface to the First Edition

The theory of branching processes is an area of mathematics that describes situations in which an entity exists for a time and then may be replaced by one, two or more entities of a similar or different type. It is a well developed and active area of research with theoretical interests and practical applications.

The theory of branching processes has made important contributions to biology and medicine since Francis Galton considered the extinction of names among the British peerage in the nineteenth century. More recently, branching processes have been successfully used to illuminate problems in the areas of molecular biology, cell biology, developmental biology, immunology, evolution, ecology, medicine, and others. For the experimentalist and clinician, branching processes have helped to understand observations that seem counter-intuitive, has helped develop new experiments and clinical protocols, and has provided predictions which have been tested in real life situations. For the mathematician, the challenge of understanding new biological and clinical observations has motivated the development of new mathematics in the field of branching processes.

The authors of this monograph are a mathematician and a cell biologist who have collaborated on investigations in the field of branching processes for more than a decade. In this monograph, we have collected examples of applications of branching processes from our own publications, and from publications of many other investigators. Each example is discussed in the context of the relevant mathematics. We have made an effort to collect and review much of the published literature which has applied branching processes to problems in molecular and cellular biology, as well as selected examples from the fields of human evolution and medicine.

The intended audiences for this monograph are mathematicians and statisticians who have had an introduction to stochastic processes but have forgotten much of their college biology, and biologists who wish to collaborate with mathematicians and statisticians. Both audiences will find many examples of successful applications of branching processes to biological and medical problems. As an aid to understand the specific examples, we have provided two introductory chapters one with background material in mathematics, and the other with background material in biology, as well as two glossaries.

The book is organized as follows: Chapter 1 provides a mathematical background and motivating examples of branching processes. Chapter 2 provides an introduction to biological terms and concepts. The subsequent chapters are divided into specific areas of branching processes. Each of these chapters develops the appropriate mathematics and discusses several applications from the published literature. Chapter 3 discusses the Galton-Watson process, the oldest, simplest and best known branching process. Chapter 4 discusses the age dependent process—Markov case, the time continuous branching process with exponential life-time distributions. Chapter 5 discusses the Bellman–Harris process, an age-dependent process. Chapter 6 gives a more systematic treatment of multitype processes, in which progeny may be of many types. Chapter 7 discusses branching processes with infinitely many types, stressing interesting properties which are different from the finite multitype situation. Appendices provide information on probability generating functions, construction of the probability space for the Bellman–Harris process as well as a brief introduction to the Jagers–Crump–Mode process (the general branching process).

We have made an effort to broadly review the published literature on branching processes applied to biology. However, we had to select specific examples and we wish to apologize to our colleagues whose work has not been cited. We welcome comments from colleagues and students who are interested the field of branching processes.

A search of any university library or an internet bookstore will reveal a number of volumes devoted to branching processes. Among the most important, we may cite the fundamental books by Harris (1963), and by Athreya and Ney (2004). Multitype branching processes were first covered in the book by Mode (1971). General branching processes, in a systematic way, were explored by Jagers (1975). Each of these classics, particularly Jagers (1975) includes some biological applications. An important book concerning estimation of branching processes is Guttorp (1991). Asmussen and Herring (1983) involve a very mathematical approach. In addition, there exist at least a dozen or two of collections of papers and more specialized volumes. Recently, Pakes (2000) prepared a report on biological applications of branching processes, which is wider in scope (it has a lot spatial branching and ecology, for example), but less detailed, although an area of overlap with our book exists. We believe that the scope of the present volume is unique in that it illustrates a paradigm, in which theoretical results are stimulated by biological applications and biological processes are illuminated by mathematics.

We gratefully acknowledge support from the following sources of support: National Institutes of Health, National Science Foundation, Keck's Center for Computational Biology at Rice University, New Jersey Commission on Cancer Research, Cancer Institute of New Jersey, Peterson Fund, Hyde and Watson Foundation, Glazer Family Fund, Rice University, Silesian Technical University and Rutgers University. Marek Kimmel was working on the final draft of this book while on a sabbatical leave at the Human Genetics Center at the University of Texas in Houston.

We thank Drs. William Sofer and Navin Sinha and several anonymous reviewers for helpful suggestions on the manuscript. Dr. Adam Bobrowski proofread the book

for its mathematical correctness. His critical remarks improved it significantly. Remaining inaccuracies are our fault. Generations of graduate students at the Statistics Department at Rice University provided welcome feedback. Professor Jim Thompson encouraged teaching this material at Rice and Rice and provided much constructive criticism. Professor Peter Jagers of the Chalmers University in Gotheborg, Sweden, hosted Marek Kimmel on several occasions and provided much needed feedback. Our families showed warmth and patience during the gestation of this book.

We dedicate this book to our students, our teachers, and our families.

Department of Statistics
Rice University
Marek Kimmel

Department of Genetics
Rutgers University
David E. Axelrod

Preface to the Second Edition

The first edition of "Branching Processes in Biology" published in 2002 has been well received by mathematicians, statisticians, and biologists. Both established investigators and advanced students have indicated to us that its inclusion of both theory and applications has been informative.

This second expanded edition adds new material published during the last decade. In addition to the work that the authors were aware of, an extensive search of the mathematical literature covered by the MatSciNet database and the biomedical literature covered by the Medline database was surveyed and relevant publications were selected. Nearly 200 new references have been added. These have been reviewed either as subsections within existing chapters, or as a new chapter.

Chapter 2, Biological Background, includes additional material on Cell Growth, Division and Death, Stem Cells, and Tumor Progression. The subsection on Textbooks and Monographs in Biology has been updated.

Chapter 3, The Galton-Watson Process, includes a new application section on Cancer Mutations, and subsections on Modeling Driver and Passenger Mutations, and a subsection on Distribution of Mutational Events in Various Phases of Tumor Growth.

Chapter 4, The Age-Dependent Process: The Markov Case. Mostly left unchanged.

Chapter 5, The Bellman-Harris Process, includes new subsections on Cell Proliferation, and on Branching Processes and Cancer Therapy.

Chapter 6, Multitype Processes, includes new subsections on Robust Modified Median Estimator of Mutation Rates, Robust Modified Median Estimator Versus Data, and Recent Developments in Theory and Application of Fluctuation Analysis.

Chapter 7, Branching Processes with Infinitely Many Types, includes a new section on Generalized Linear-Fractional Distributions and Their Applications, with subsections on Definitions and Basic Properties, and Applications in Branching Processes. A new section Application of Branching Process with Infinite-Allele Mutations includes subsections on Proliferation of Alu Repeats, and Modeling Telomeres.

Chapter 8, Genealogies of Branching Processes and Their Applications, is a new chapter, with a subsection on Robustness of Mitochondrial Eve Dating.

Chapter 9, References. Nearly 200 new recent references have been added, bringing the total to over 460 references.

Chapter D, Glossaries. New entries have been added to define or explain new terms.

In addition to the revised subject index, a new author index has been added.

New applications have been added in appropriate chapters. They discuss recent advances in several areas, including cancer mutations, cancer therapy, cell proliferation, estimation of mutation rates, fluctuation analysis, *Alu* repeats, and telomeres, among others.

During the past decade two of our colleagues, who had made major contributions to this field, have died, Ovide Arino and Andrey Yakovlev. We hope that their work, and the other work reviewed in this edition, will inspire new investigators to explore this active field of branching processes in biology.

We thank Tomasz Wojdyla for preparing the manuscript, and Thomas McDonald and Nicolas Flores for reading and suggesting changes.

Guide to Applications, or How to Read This Book

As mentioned in the Preface to the First Edition, the book is organized by different classes of branching processes, except for Chap. 1, providing general motivation and some mathematical background and Chap. 2, providing biological background. Two glossaries at the end of the book give definitions of basic biological and mathematical terms commonly used in the book. The inner structure of the book is a network of interconnected biological applications, which increase in detail when modeled by progressively more sophisticated branching processes. The list below includes major application sections, some of them based upon our own work.

- Cancer Chemotherapy
  - Analysis of the stathmokinetic experiment, Sect. 5.4
  - Stochastic model of cell cycle with chemotherapy, Sect. 6.4
- Cell cycle models
  - Simplest version with death and quiescence, Sect. 3.2
  - Unequal division and growth regulation, Sect. 7.8.1
  - A model of two cell populations, Sect. 6.3
  - Structured cell population models, Sect. 7.8
- Evolution theory
  - Complexity threshold in early life, Sect. 3.4
  - Galton-Watson processes in random environment and macroevolution, Sect. 3.9
  - Age of mitochondrial Eve, Sect. 8.3
  - Yule's evolutionary model, Sect. 7.9
- Gene amplification and loss of telomere sequences
  - Galton-Watson branching process model, Sect. 3.7
  - A model of unstable gene amplification, Sect. 7.4
  - Stable gene amplification, Sect. 7.5
  - Quasistationarity in a branching model of division-within-division, Sect. 7.6
  - Loss of telomere sequences, Sect. 7.7
- Molecular biology
  - Cell surface aggregation phenomena, Sect. 6.5
  - Polymerase chain reaction, Sects. 1.2 and 6.8

- Mutations
  - Cancer mutations, Sect. 3.6
  - Iterated Galton-Watson process and expansion of DNA repeats, Sect. 3.8
  - Clonal resistance theory, Sect. 4.2
  - Mutations and fluctuation analysis, Sect. 6.1
  - Deletions in mitochondrial DNA, Sect. 6.7
  - Branching process with infinite-allele mutations, Sect. 7.10

In addition, there are shorter application-related sections spread over the book, mostly containing examples taken from the literature. These include diverse applications of the Galton–Watson process in Sect. 3.10, examples of branching process models of cell proliferation and estimation of cell cycle parameters in Sect. 5.5, and infinite-type branching process models in cell biology, genetics and cancer in Sect. 6.9.

Contents

Chapter 1
Motivating Examples and Other Preliminaries

The branching process is a system of particles (individuals, cells, molecules, etc.) which live for a random time and, at some point during lifetime or at the moment of death, produce a random number of progenies. Processes allowing production of new individuals during a parent individual's lifetime are called the general or Jagers–Crump–Mode processes (Fig. 1.1a). They are suitable for description of populations of higher organisms like vertebrates and plants. Processes that assume production of progeny at the terminal point of parent entity's lifetime are called the classical processes (Fig. 1.1b). They are usually sufficient for modeling populations of biological cells, genes, or biomolecules. In some processes, like the time-continuous Markov process, the distinction is immaterial since one of the progeny of a particle may be considered an extension of the parent.

One of the important notions in the theory of branching processes is that of the type space. The type space is the set, which can be unite, denumerable, or a continuum, of all possible varieties of particles included in the process. Particles of a given type may produce particles of different types. Restrictions on type transitions, as well as on the type space, lead to differing properties of resulting processes. The richness of these classifications is already apparent on the level of denumerable type spaces.

1.1 Some Motivating Examples

One of the oldest branching processes ever considered was the process in which "particles" were male individuals bearing noble English family names. An ancestor in such a process initiated a pedigree which might inevitably become extinct if all the male descendants died without heirs. Is extinction of a noble family name inevitable in the long run? How many generations will elapse before extinction occurs? These are typical questions asked about a process in which the number of progeny of an individual may be equal to zero. Similar questions may be posed in a situation when a mutant cell initiates a small colony of precursor cancer cells. How likely are these colonies to die before they become numerous enough to overgrow the surrounding normal type?

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_1

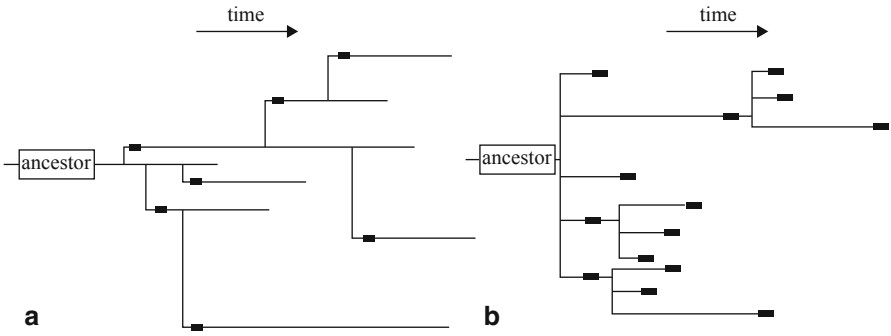


Fig. 1.1 General (*panel a*) and classical (*panel b*) branching processes. *Black rectangles* depict individuals (objects, particles, etc.), *horizontal lines* depict lifetimes. *Vertical lines* are added to link individuals to their parents. The length of *vertical lines* is arbitrary

A different type of question may be posed for processes in which the growth is assured by a sufficiently high proliferation rate. Then, the interesting parameter is the long-term growth rate and the size and composition of the population at a given time. This is typical of laboratory populations of biological cells cultured with abundant nutrients and sufficient space. The same is true of prosperous individuals settling in a large territory with little obstacles to growth. An interesting example is provided in the book by Demos (1982), where it is stated that the average number of progeny surviving to maturity among the British colonists in New England in the seventeenth century was equal to nine.

The patterns of branching may be quite complicated. An interesting example was given in the book by Harris (1963). In the course of evolution, new species are created by successful new varieties of organisms which become reproductively separated from their ancestral species. This is an ordinary branching process. However, from time to time an event occurs, which creates a species so novel that it has to be considered an ancestor of a higher taxonomic unit than a species: A family. Therefore, branching becomes hierarchical: Small particles (species) proliferate inside of large particles (families) which proliferate themselves, each started by a founder species. At both levels, extinction may occur. A similar branching pattern describes AIDS viruses proliferating in human T-lymphocyte cells. Divided lymphocytes inherit a portion of viruses present in the parent cell. If the number of viruses in a cell exceed a threshold, the cell dies. In this example, the two levels of branching compete with each other. Still another pattern is found in cancer cells, inside which multiple copies of a gene increase a cell's resistance to treatment. If there are not enough of these, the cell becomes sensitive and dies.

1.2 Application: Polymerase Chain Reaction and Branching Processes

This section considers an important example of a branching process describing one of the most important tools of molecular biology, the polymerase chain reaction (PCR). Following an introduction, we present a mathematical and simulation model constructed by Weiss and von Haeseler (1997). Material of this section is based on Weiss and von Haeseler (1997) if not stated otherwise. Finally, we describe an application of PCR in artificial evolution.

1.2.1 Introduction About the Mechanics of PCR

The following introduction has been adapted from a thesis by Shaw (2000): The PCR is an experimental system for producing large amounts of genetic material from a small initial sample. The reaction performs repeated cycles of DNA replication in a test tube that contains free nucleotides, DNA replication enzymes, and template DNA molecules. The PCR amplification technique operates by harnessing the natural replication scheme of DNA molecules and using a naturally derived DNA polymerase protein. The result of the PCR is a vast amplification of a particular DNA locus from a small initial number of molecules.

Another feature of the PCR process is the stochasticity of amplification. Amplification is random because not every existing molecule is successfully replicated in every reaction cycle. Experimental evidence suggests that even the most highly efficient reactions operate at an efficiency around 0.8, i.e., each double-stranded molecule produces an average of 0.8 new molecules in a given reaction cycle. The randomness in PCR can be attributed to the multiple molecular events which must occur in order to copy DNA.

The purpose of the PCR process is to produce clones (subpopulation with common descent) from DNA molecules from the small initial sample. Under ideal conditions, in each clone the molecules are identical, in the sense that the sequence of nucleotides A, T, C,and G in each molecule is either identical or complementary (A, T, C, and G replaced by T, A, G, and C, respectively) to the ancestral molecule of the clone (molecules in the initial samples may not be identical).

However, random alterations of nucleotides in DNA sequences, known as mutations, also occur during PCR amplification. In many PCR applications, mutations which occur during the PCR hinder analysis of the initial sample, such as in the forensic setting. In other settings, however, PCR mutations are desirable, as is the case in site-directed mutagenesis studies and artificial evolution experiments (Joyce 1992). In both situations, analysis of variants generated during PCR is required. Interest focuses on the study of genetic diversity in a sample of molecules from the final stage of a PCR experiment. The molecules sampled are potentially related as descendants of a common ancestor molecule. The common ancestor of a family of

PCR products is an initial molecule present at the starting stage of the amplification. The sampled molecules more commonly represent k samples of size 1 from distinct ancestor particles. This situation arises commonly because PCR is performed from a very large number of initial molecules, usually more than many thousands. In either case, the genealogical method may be used to analyze the diversity of a sample taking into account the replication history and relatedness of the sampled molecules.

In order to assess the genealogy of the molecules in a sample, one must model the PCR and the structure of DNA replication. As in natural systems, DNA replication in the PCR is semiconservative, so that only one strand of each double-stranded DNA molecule is newly manufactured in a single replication event. Replication is semiconservative because each new single strand is built from a complementary antiparallel template strand during replication. Mutations can occur during construction of the new strand, so that newly fabricated strands may not be fully complementary to their templates. If a mutation occurs at some intermediate cycle of the PCR, the mutation will be propagated by the amplification procedure into all descendants of the mutant molecule. The goal is to study the sequence diversity of DNA molecules resulting from mutations during amplification.

1.2.2 Mathematical Model

A model of PCR has to include a model of the replication process and the mutation process. We use the single-stranded model, which is a simplification, because DNA is double stranded (see a discussion in Shaw 2000). In the following, we frequently use molecule as a synonym for single-stranded sequence containing the subsequence of interest. Any other chemical molecules that are, in reality, present in a PCR tube are not considered. The replication process of PCR is described in terms of branching processes. The reaction proceeds through discrete cycles involving thermal and chemical processes. In each cycle, each single-stranded template should produce a copy. So, ideally, PCR is a binary fission process with discrete time (a special case of a Galton–Watson process, see Chap. 3). We assume that a PCR starts with S_0 identical copies of single-stranded sequences. Let S_i be the number of sequences present after the ith cycle. In cycle i,each of the S_{i-1} template molecules is amplified independently of the others with probability λ_i. The probability λ_i can also be viewed as the proportion of amplified molecules in cycle i, hence, it is called the efficiency in cycle i. More precisely, the efficiency does not simply depend on the cycle number, but on the number of amplifications in the previous cycles and on PCR conditions.

If we assume that the random variable S_i depends only on λ_i and S_{i-1} then the sequence $S_0, S_1, \ldots, S_i, \ldots$ forms a nonhomogeneous binary fission. If $\lambda_i = \lambda$ is independent of the cycle number, then the accumulation of PCR product is a Galton–Watson branching process.

Because replication in PCR is not error free, we add a mutation process to the model: We assume that a new mutation occurs at a position that has not mutated in any other sequence before. Furthermore, we model the process of nucleotide substitution

as a Poisson process with parameter μ, where μ is the error rate (mutation rate) of PCR per target sequence and per replication. This so-called infinitely many sites model (ISM) does not allow for parallel or back mutations. In the case of PCR, this assumption seems reasonable because only a small number of mutations are observed in practice.

1.2.3 Genealogical Approach

Computer simulations of stochastic processes have become a powerful tool to analyze data in situations where analytical methods are not feasible. In the population genetics literature, a prominent example is the coalescence process that describes the ancestral relationship between a sample of individuals in a population as one goes back in time (Tavaré 1984). Rather than trying to analyze the relations of all individuals in a population, the coalescence describes the (unknown) genealogy of a sample in terms of a stochastic process. If one starts with a sample of size n and traces back the ancestral history of these n lineages, one can compute the probability that at a time t, two lineages in the genealogy coalesce, i.e., the most recent common ancestor (MRCA) of the corresponding individuals is found. The probability depends on the sample size and the population trajectory. After a coalescent event occurs, the number of lineages is reduced by 1. The coalescent process stops when the MRCA of the whole sample is found. In the situation of a stationary population of constant size, simple formulae are available that describe branch lengths in a genealogy of a sample, total length of a genealogy, etc. If one drops the assumption of constant size, it is more difficult to find analytical solutions, whereas it is still possible to get instructive results using simulation techniques.

The following simulation method to analyze PCR data bears similarity with the coalescent approach (see Sect. 8.1 for a more mathematical treatment of a similar process): In PCR, the offspring of the initial molecules are related by a randomly growing tree. Instead of generating this tree independently for each initial template, we adopt the following approach: For each initial molecule the number of molecules in the PCR product in each cycle (the size trajectory) is computed (step 1 in the algorithm). Thereby, we distinguish two types of molecules: those that are immediate copies from a molecule of a previous cycle (filled circles in Fig. 1.2) and those that existed in the previous cycle (open circles in Fig. 1.2). From all molecules at the end of the PCR a random sample of n sequences is drawn. Then, we randomly assign to each of the sampled sequences one of the initial molecules as ancestor (step 2) and regard the sets of sampled sequences that are descendents of the same initial molecule as subsamples. In the next step (step 3), we trace back the genealogies of all subsamples separately.

Figure 1.2 illustrates this process for one initial molecule (and one subsample). In this example, we assume that a subsample of 6 sequences was drawn from a total of 16 sequences. In order to generate the subsample genealogy, the special features of PCR must be taken into account. A coalescent event, i.e., the merging of two

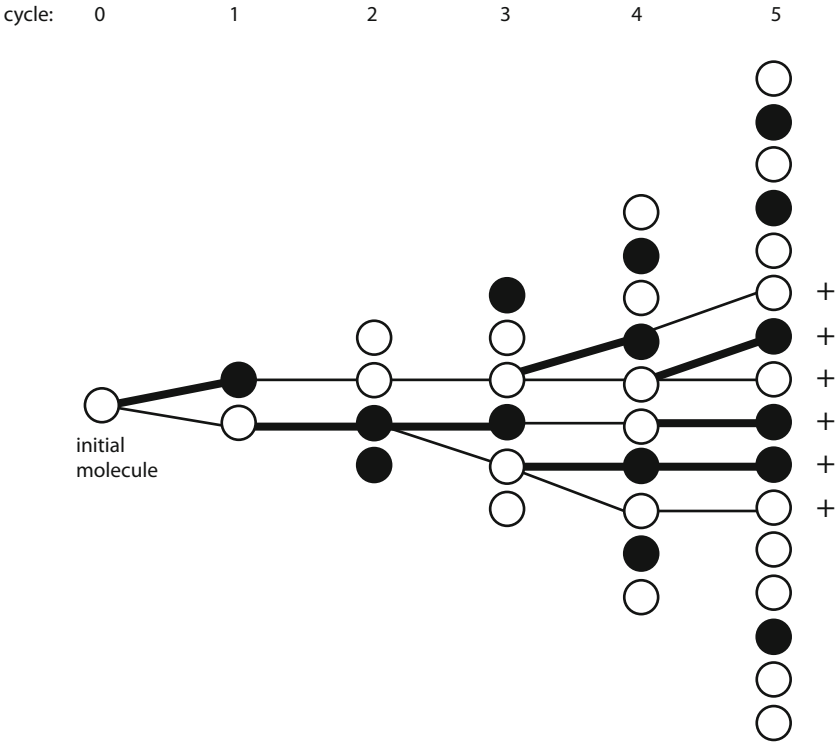


Fig. 1.2 Graphical illustration of a subsample genealogy according to the example considered. *Filled circles* represent molecules that were newly amplified in a cycle, *open circles* represent molecules already present in the previous cycle, + indicates that the molecule is in the sample, *thick lines* represent a replication in the genealogy. (Source: Weiss and von Haeseler 1997)

molecules while going from cycle 5 to cycle 4, is only possible if exactly one of the two molecules is an immediate copy. Among the six sampled sequences, three were copied during cycle 5. Hence, at the most, three coalescent events are possible, and in fact one such event occurred. The coalescent process stops when only one molecule is left. If only one molecule is present and the cycle number is not equal to zero, we have to determine how many replications took place from the initial molecule to this molecule.

After all subsample genealogies are generated, they are combined to one single genealogy (step 4). Finally, we superimpose a mutational process on the genealogy (step 5), where mutations are only allowed where replications took place in the genealogy (thick lines in Fig. 1.2). Before we describe the simulation more formally, we assign a number $k, k = 1, \ldots, S_0$ to each of the S_0 initial molecules.

1.2.4 Statistical Estimation of the Mutation Rate

Weiss and von Haeseler (1997) carried out estimation of the mutation rate μ, based on a published data set. They used a convenient measure of the diversity of the sample resulting from PCR errors (mutations), defined as the number M_n of variable positions in a sample of size n, i.e., the number of the entries of the DNA sequence at which two or more variants are observed in the sample. In genetic literature, M_n is also known as the number of segregating sites.

Weiss and von Haeseler (1997) used the data of Saiki et al. (1988) who amplified a 239-base-pair region (i.e., a DNA sequence 239 nucleotides long) of genomic DNA. After $C = 30$ PCR cycles, $M_{28} = 17$ variable positions were observed when they sequenced $n = 28$ different clones. Furthermore, the authors measured the extent of amplification after 20, 25, and 30 cycles. They report an increase of 2.8×10^5, 4.5×10^6, and 8.9×10^6, respectively. This corresponds to an overall efficiency of 0.705 in 30 cycles. They also determined cycle-specific efficiencies from the data using the following formula:

$$\frac{\mathrm{E}(S_i)}{\mathrm{E}(S_{i-j})} = (1 + \lambda_i)^j, i \geq j.$$

From the reported increase after 20, 25, and 30 cycles they computed:

$$\lambda_i = \begin{cases} 0.872, & i = 1, \ldots, 20, \\ 0.743, & i = 21, \ldots, 25, \\ 0.146, & i = 26, \ldots, 30. \end{cases}$$

These values for λ_i were used in the simulations. Since no information about the number of initial molecules is given, the analysis was carried out for different S_0 values (1, 10, 100, 1000). They generated probability distributions $\Pr(M_n = m|\mu)$ for 100 equidistant values of μ in the interval $[0.019, 0.079]$. The scale on the m axis is limited to 30, since for the range of μ values considered, the likelihood has very small values for $m > 30$. If one takes the observed number of variable positions in the sample equal to 17 and cuts along this line through Fig. 1.3, one gets $lik(\mu|M_{28} = 17)$, the likelihood function of μ given $M_{28} = 17$. Figure 1.4 shows the likelihood functions for $S_0 = 1, 10, 100$, and 1000. For each S_0, the position of the maximum of the likelihood function is the maximum likelihood estimate of μ.

1.2.5 Mutagenic PCR and Artificial Evolution

As mentioned above, mutations in the PCR may be desirable. One such example is provided by artificial evolution experiments, in which biomolecules, like RNA enzymes (ribozymes), are subjected to alternating rounds of amplification and mutation, and selection. In some classical experiments (Joyce 1992, Beaudry and Joyce 1992),

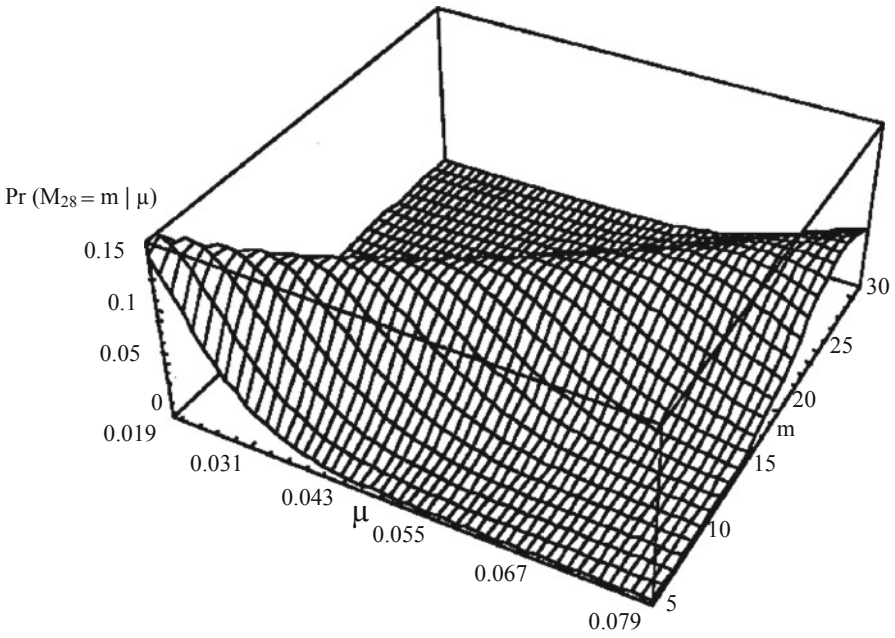


Fig. 1.3 Example of simulated probability distributions $\Pr(M_n = m|\mu)$ for 100 equidistant values of μ ($S_0 = 100$). (Source: Weiss and von Haeseler 1997)

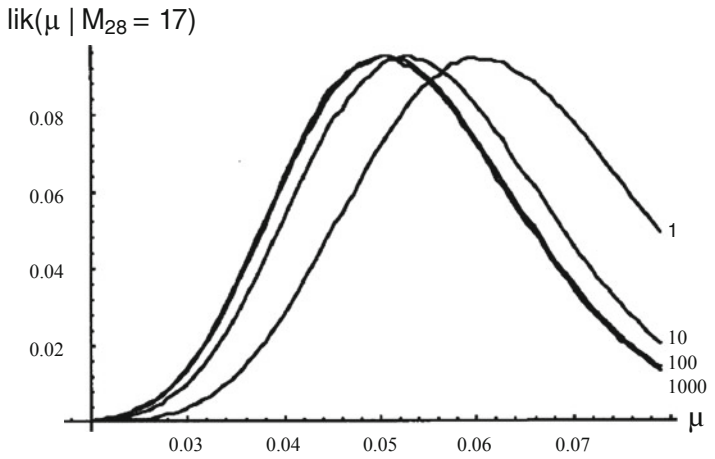


Fig. 1.4 Simulated likelihood functions $lik(\mu|M_{28} = 17)$ for a published data set. The numbers in the graph represent the used S_0 values. (Source: Weiss and von Haeseler 1997)

high functional specificity of the evolved product was achieved. In these experiments, mutations provide substrate for selection and therefore understanding the mutational process in these experiments is very important. As of now, this remains an open problem.

1.3 The Branching Property

The branching property is a basic feature identifying processes studied in this book. It is responsible for many properties of the branching processes, some of them unexpected. The basic assumption involved is that each particle in the process behaves identically as all other particles and independently of all other particles (this latter, conditionally on its existence). This may appear simple and obvious. However, consequences are far reaching. Let us consider the clone, extending indefinitely into the future, originating from an ancestral particle. Such a clone is identical to the entire process we are studying. If we take any particle from this clone, then it gives rise to its own clone, which is a subprocess of the whole process. However, by the branching property, this subprocess is identical to the whole process. This realization provides a way to describe the process mathematically: It can be decomposed into subprocesses, which are identical (identically distributed, to be rigorous) to each other and to the entire process. In mathematical terms, branching processes belong to a class of stochastic objects called "self-recurrent" by Feller (1968, 1971).

Matters become a little more complicated if we allow particles of different types. The clones created by particles of different types are different, so the bookkeeping becomes more involved. However, the principle stays the same. The rest of this section is concerned with mathematical notation and it can be safely skipped at first reading.

Let us consider a classical branching process in which progeny is born at the moment of parent's death. It can be understood as a family $\{Z(t,\omega), t \geq 0\}$ of nonnegative integer-valued random variables defined on a common probability space Ω with elements ω. The branching process is initiated at time $t = 0$ by the birth of a single ancestor particle. Suppose that the life length of the ancestor is a random variable $\tau(\omega)$ and that the number of its progeny (produced at its death) is equal to $X(\omega)$ (Fig. 1.5). Each of the progeny can be treated as the ancestor of its own process, which is a component of our branching process. Then, the number of individuals present in the process at time t is equal to the sum of numbers of the individuals present in all these subprocesses. This bookkeeping is correct for $t \geq \tau(\omega)$, i.e., after the ancestor has died. Before the ancestor's death, the number of particles is equal to 1. Summarizing:

$$Z(t,\omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^{(i)}(t, \tau(\omega), \omega), & t \geq \tau(\omega), \\ 1, & t < \tau(\omega), \end{cases} \tag{1.1}$$

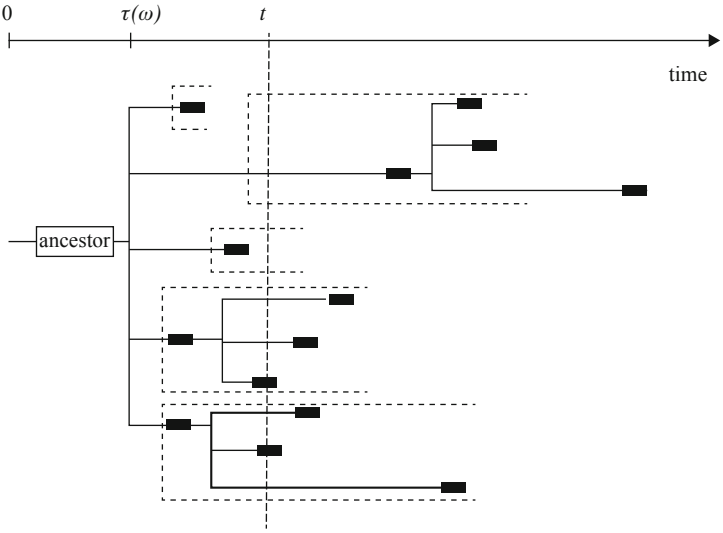


Fig. 1.5 Decomposition of the branching process into subprocesses generated by the first-generation progeny of the ancestor, see Eq. (1.1). In the case depicted, the number of the first-generation progeny is equal to $X(\omega) = 5$. At time $t > \tau(\omega)$, the number of particles in the subprocesses generated by progeny 1, 2, 3, 4, and 5, is equal, respectively, to $Z^{(1)}(t, \tau(\omega), \omega) = 0$, $Z^{(2)}(t, \tau(\omega), \omega) = 1$, $Z^{(3)}(t, \tau(\omega), \omega) = 0$, $Z^{(4)}(t, \tau(\omega), \omega) = 3$, and $Z^{(5)}(t, \tau(\omega), \omega) = 3$. The total number of particles in the process at time t is equal to 7

where $Z(t, \tau(\omega), \omega)$ denotes the number of individuals time t in the process started by a single ancestor born at time $\tau(\omega)$, and the additional superscript (i) denotes the ith independent identically distributed (iid) copy. Double summation is needed since ancestor's progeny generally may be of all possible types. The self-recurrence (or branching) property is embodied in the statement that the processes initiated by the progeny of the ancestor are independent and distributed identically as the ancestor,

$$Z^{(i)}(t, \tau(\omega), \omega) \stackrel{d}{=} Z^{(i)}(t - \tau(\omega), \omega). \tag{1.2}$$

Substitution of expression (1.2) into (1.1) leads to a recurrent relation

$$Z(t, \omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^{(i)}(t - \tau(\omega), \omega), & t \geq \tau(\omega), \\ 1, & t < \tau(\omega), \end{cases}$$

which we will use repeatedly.

In a rigorous way, the construction of a branching process proceeds from specification of distributions of life lengths and progeny numbers of individuals to the construction of the probability space Ω to deriving a specific form of relationships (1.1) and (1.2) . Based on a classical construction by Harris (1963), the procedure has been extended to most general processes. In our applications, the existence and

form of the probability space and self-recurrent relationships of the type (1.1)–(1.2) will be obvious. Therefore, usually, we will drop from the notation the argument ω, although implicitly it is always present.

The self-recurrence characterizing the branching process is one of the two conceivable ways the process can be defined. It is based on decomposing the process into a union of subprocesses initiated by the direct descendents of the ancestor. It can be called the "backward" approach, in an analogy to the backward Chapman–Kolmogorov equations of Markov processes. A dual "forward" approach consists of freezing the process at time t, recording the states of all individuals at that time, and predicting their future paths (e.g., at $t+1$ or at $t+\delta t$). The backward–forward duality will be useful in some of our considerations.

Branching processes have been widely used to describe growth and decay of biological populations. Their use has always overlapped with that of deterministic mathematical tools like ordinary and partial differential equations. The doubtless applicability of branching processes is in studying small populations in which random fluctuations play a major role. However, some results concerning large populations are also easier to deduce using branching processes (see, e.g., Arino and Kimmel 1993).

1.4 Probability Generating Functions and Analytical Methods

Consider a branching process composed of particles of one type. The number of particles at time t is denoted $Z(t)$. An ancestor is born at $t=0$ and at random time τ it gives birth to a random count of progeny. Each of the progeny initiates a subprocess identical to the whole process. Therefore, conditional on τ,

$$Z(t) \stackrel{d}{=} \sum_{i=1}^{X} Z^{(i)}(t-\tau), t \geq \tau, \tag{1.3}$$

where $Z^{(i)}(t)$ is the number of particles in the ith independent copy of the process. Therefore, $Z(t)$ can be represented as a sum of a random number of iid random variables (rv), with nonnegative integer values. A useful tool for handling distributions of such random sums is the probability generating function (pgf) of a distribution. Methods employing pgf manipulations instead of directly dealing with random variables are called analytic.

Pgf are the basic analytic tool employed to deal with nonnegative rv's and finite and denumerable sequences (vectors) of such variables. Let us denote Z_+ as the set of nonnegative integers. Let X be a Z_+-valued rv, such that $\mathrm{P}[X=i]=p_i$. We write $X \sim \{p_i\}_{i\geq 0}$ and say that $\{p_i\}$ is the distribution of X.

Definition 1.1 (Definition of the pgf). *The pgf f_X of a Z_+-valued rv X is a function $f_X(s) = E(s^X) = \sum_{i=0}^{\infty} p_i s^i$, of a symbolic argument $s \in U \equiv [0,1]$. With some abuse of notation, we write $X \sim f_X(s)$. The following are the basic properties of the pgfs.*

The following theorem is a collection of results, which usually are given separately. All can be found, e.g., in the book by Feller (1968).

Theorem 1.1 Pgf theorem. *Suppose X is a Z_+-valued rv with pgf $f_X(s)$ which may not be proper. Let us denote (N) the nontriviality condition $p_0 + p_1 < 1$.*

1 *f_X is nonnegative and continuous with all derivatives on $[0, 1)$. Under (N), f_X is increasing and convex.*
2 *If X is proper, $f_X(1) = 1$,otherwise $f_X(1) = P[X < \infty]$.*
3 *$d^k f_X(0)/ds^k = k!p_k$.*
4 *If X is proper, the kth factorial moment of X, $\mu_k = E[X(X-1)(X-1)\cdots(X-k+1)]$, is finite if and only if $f_X^{(k)}(1-) = \lim_{s\uparrow 1} f_X^{(k)}(s)$ is finite. In such case, $\mu_k = f_X^{(k)}(1-)$.*
5 *If X and Y are two independent Z_+-valued rv's, then $f_{X+Y}(s) = f_X(s) f_Y(s)$.*
6 *If Y is a Z_+-valued rv and $\{X^{(i)},\ i \geq 1\}$ is a sequence of iid Z_+-valued rv's independent of Y, then $V = \sum_{i=1}^{Y} X^{(i)}$ has pgf $f_V(s) = f_Y[f_{X^{(1)}}(s)]$.*
7 *Suppose that $\{X_i,\ i \geq 1\}$ is a sequence of Z_+-valued rv's. $\lim_{i\to\infty} f_{X_i}(s) = f_X(s)$ exists for each $s \in [0, 1)$ if and only if the sequence $\{X_i,\ i \geq 1\}$ converges in distribution to a rv X, i.e., if limits $\lim_{i\to\infty} P[X_i = k]$ exist for all k and are equal to $P[X_i = k]$, respectively. Then $f_X(s)$ is the pgf of the limit rv X.*

The definition of the pgf can be generalized to the multivariate and denumerable cases (Appendix).

Returning to the example in the beginning of this section (Eq. 1.3), we notice that based on the pgf theorem, part 6, it can be now replaced by an equivalent pgf identity:

$$f_t(s) = f[f_{t-\tau}(s)],\ t \geq \tau,$$

conditional on τ (i.e., given the ancestor dies at age τ), where $f_t(s)$ denotes the pgf of Z_t and $f(s)$ the pgf of X. As an example, let us consider the Galton–Watson process, which will be studied in detail in Chap. 3. In this process, all particles including the ancestor live for a fixed time equal to 1, so that $\tau \equiv 1$. This means that $f_t(s) = f\,[f_{t-1}(s)]$ for all $t \geq 1$, and so $f_t(s) = \underbrace{f\,\{f\,[\cdots f_{t-n}}_{n}(s)\cdots]\}$. If we limit ourselves to integer t and notice that $f_0(s) = s$, i.e., at $t = 0$ only the ancestor is present, then we obtain

$$f_t(s) = \underbrace{f\,\{f\,[\cdots f}_{t}(s)\cdots]\}, \tag{1.4}$$

which is the pgf law of evolution of the Galton–Watson process.

1.5 Classifications of the Branching Processes

1.5.1 Lifetime

The distribution of particle lifetime τ has much impact on the behavior and analysis of the process. As shown above, if $\tau \equiv 1$ (the Galton–Watson process, Chap. 3), it is enough to consider integer times. The pgf of Z_t (Z_t is an accepted notation for $Z(t)$ when time is discrete) is simply the t-fold functional iterate of the pgf of the progeny number, X (Eq. 1.4).

Another important special case is when τ is distributed exponentially. The lack of memory of the exponential distribution leads to a process with continuous time which can be considered an interpolation of the Galton–Watson process between integer time points (Chap. 4).

Finally, if τ is an arbitrary nonnegative random variable, the resulting process is called "age-dependent" or Bellman–Harris process. It is more complicated to analyze than any of the two previous processes (Chap. 5).

1.5.2 Type Space

The following is the list of the frequent variants of type space:

$$S = \begin{cases} \{1\}, & \text{single type,} \\ \{1, \dots k\}, & \text{multitype,} \\ \{1, 2 \dots\}, & \text{denumerable type,} \\ R_+, R, [0,1], & \text{continuous type,} \\ \text{abstract.} & \end{cases}$$

The Galton–Watson and Bellman–Harris processes considered above are single type but have their multitype, denumerable-type and continuous-type counterparts (Chaps. 6 and 7). Abstract type spaces are used to create "superindividuals" composed of a number of original individuals and in this way to handle dependence among particles (Taïb 1997)

1.5.3 Criticality

A very important classification is based on the mean progeny count $m = \mathrm{E}(X)$ of a particle. We introduce it using the example of the Galton–Watson process, but it is valid for all branching processes. By the pgf theorem, $\mathrm{E}(X) = f'(1-)$ and $\mathrm{E}(Z_t) = f_t'(1-)$. Differentiating the formula $f_t(s) = \underbrace{f\,\{f\,[\cdots f}_{t}(s)\cdots]\}$ with

respect to s and substituting $s = 1$, we obtain using the chain rule of differentiation,

$$\mathrm{E}[Z_t] = f_t'(1-) = [f'(1-)]^t = m^t.$$

Therefore, in the expected value sense, the process grows geometrically if $m > 1$, stays constant if $m = 1$, and decays geometrically if $m < 1$. These three cases are called supercritical, critical, and subcritical, respectively:

$$\begin{cases} m > 1, & \text{supercritical} & \Rightarrow & \mathrm{E}[Z_t] \uparrow \infty, \\ m = 1, & \text{critical} & \Rightarrow & \mathrm{E}[Z_t] = 1, \\ m < 1, & \text{subcritical} & \Rightarrow & \mathrm{E}[Z_t] \downarrow 0. \end{cases} \tag{1.5}$$

The above relationships are intuitively expected. However, the corresponding laws of extinction are less intuitive. Let us consider the probability $q_t = f_t(0) = \mathrm{P}[Z_t = 0]$ that the process is extinct at time t. We have $q_{t+1} \geq q_t$, since $Z_t = 0$ implies $Z_{t+1} = 0$. Since also $0 \leq q_t \leq 1$, the sequence $\{q_t\}$ tends to a limit q which is the probability of eventual extinction. Moreover, since $f_{t+1}(s) = f\,[f_t(s)]$ then, setting $s = 0$, we obtain $q_{t+1} = f(q_t)$ and, letting $t \to \infty$, $q = f(q)$. Therefore, q is the coordinate at which $f(s)$ intersects the diagonal. Let us notice that $f(s)$ is convex and $f(1) = 1$. If $m = f'(1-) > 1$, then there exists $0 \leq q < 1$ such that $f(q) = q$. If $m = f'(1-) \leq 1$, then q has to be equal to 1. Therefore we obtain that

$$\begin{cases} m > 1, & \text{supercritical} \\ m = 1, & \text{critical} \\ m < 1, & \text{subcritical} \end{cases} \quad \begin{array}{l} \Rightarrow \quad q < 1, \\ \\ \Rightarrow \quad q = 1. \end{array} \tag{1.6}$$

The supercritical and subcritical processes behave as expected from the expression for the means. The critical process is counterintuitive. Although the mean value stays constant and equal to 1, the process becomes extinct almost surely ($q = 1$). This latter is only possible if the tail of the distribution is heavy enough to counterbalance the atom at 0. This suggests that a critical process is undergoing large fluctuations before it becomes extinct [c.f., the discussion following Eq. (3.7)].

Further on, in Chaps. 3–5 we will see that the limit behavior in all three cases may be characterized in more detail.

Critical processes have been widely considered as models of self-organization. One example is the paper by Kim et al. (2009). It has been known that many complex networks in real world are fractals, satisfying the fractal scaling: the number of boxes (NB) needed to cover an object scales in a power-law manner with respect to the box size. Examples are the World Wide Web, the protein interaction network of budding yeast, and the metabolic networks. In contrast, many artificial model networks such as the Barábasi-Albert (BA) model, are not fractals. Random critical branching trees (CBTs) are generated by the multiplicative branching process, where the branching number is determined stochastically, independent of the degree of their ancestor.

The authors demonstrate analytically that despite this stochastic independence, there exists the degree–degree correlation (DDC) in the CBT and it is disassortative. Moreover, the skeletons of fractal networks, the maximum spanning trees formed by the edge betweenness centrality, behave similarly to the CBT in the DDC. This analytic solution and observation support the argument that the fractal scaling in complex networks originates from the disassortativity in the DDC, which has been pondered for a number of years. Further details exceed the scope of the book.

1.6 Modeling with Branching Processes

In this section, we discuss branching processes as a modeling tool, in a general and philosophical way. Our discussion owes a lot to ideas presented in a review paper by Jagers (1991). We complement it with our own insights concerning the interactions between biology and branching processes.

As stated by Jagers (1991), "Mathematical population theory is not the same as demography: Its object of study is not human populations. Nor is its object actual biological populations of, say, animals, bacteria or cells, or the physical populations of splitting particles in a cascade or neutron transport. Rather, its purpose is to study the common theme of these and many other empirical phenomena, an idealized pattern of free population growth, of sets changing as their members generate new set members."

For a mathematician, "The essence of such a theory is mathematical in the same sense as geometry, the study of idealized shape. It is relevant for actual populations in so far as their reproduction is close to the idealized free reproduction and to the extent that this reproduction property is important for the evolution of the system as a whole. Thus in vitro cell kinetics is close to the pattern, at least if the population has enough nutrition and space, whereas the well-regulated growth of a couple of fetus cells into, say, a hand is dominated by features other than population growth."

"The population growth pattern is an important one, often playing a great role in the evolution of phenomena, and it can be discerned in many circumstances, ranging not only from demography to particle physics but including even data structures for sorting and searching in computer science or fractal sets arising in various types of mathematics. Sometimes the conclusions you can draw from the general mathematical study are even stronger than those obtained through more specialized models."

The approach advocated in our book, is to explain biological observations in detail, the way mathematics is used in theoretical mechanics or relativity, and to generate predictions accurate enough to be practical. This approach may be considered a type of engineering. One may argue that, by doing so, the modelers enter the turf reserved for other professions: biostatistics, demography, computer simulation, and biotechnology. Still, mathematical principles alone can explain the balance of factors contributing to the behavior of a population as a whole.

Unfortunately, this is not always appreciated in biology. One of the reasons is that much of modern biology is molecular biology. This latter, through introduction of

new techniques for gathering data and probing biological processes at a fundamental level, continuously provides an unprecedented amount of new information. Much of this information is connected only at a simplistic level (Maddox 1992). In the extreme reductionist view everything can be reduced to a molecular switch which turns on or cuts off expression of a gene. In reality, it is frequently a delicate dynamic balance that creates a given behavior, and there might be alternative ways of inducing a biological system to display a seemingly related set of properties. For example, a complex human genetic disease, like diabetes, can arise through many alternative molecular pathways. A book by a well-known evolutionist discusses this subject (Lewontin 2000).

However, there are reasons to think that this situation soon may change. Mathematical and computational methods are steadily making new contributions to molecular biology. One example is the progress achieved in analysis of DNA sequences using Hidden Markov models (Durbin et al. 1998). Beginning with the sequencing of the human genome (Venter et al. 2001; International Human Genome Sequencing Consortium 2001), a flood of data have been generated concerning DNA sequences and their expression. This information is having an impact on our understanding of many areas of biology and medicine, including evolution and human diseases. Resulting problems will be difficult to resolve without mathematics and/or computing power.

The subject of this book is the use of branching processes to model biological phenomena of some complexity, at different, though predominantly cellular or subcellular, levels. To understand the power and the limitations of this methodology, again we follow Jagers (1991). Probabilistic population dynamics arises from the interplay of the population growth pattern with probability. Thus, the classical Galton–Watson branching process defines the pattern of population growth using sums of iid rv; the population evolves from generation to generation by the individuals getting iid numbers of children. This mode of proliferation is frequently referred to as "free growth" or "free reproduction".

The formalism of the Galton–Watson process provides insight into one of the fundamental problems of actual populations, the extinction problem and its complement, the question of size stabilization: If a freely reproducing population does not die out, can it stabilize, or must it grow beyond bounds? The answer is that there are no freely reproducing populations with stable sizes (see Sect. 3.3 for mathematical details). Population size stability, if it exists in the real world, is the result of forces other than individual reproduction, of the interplay between populations and their environment. This is true for structures much more general than the Galton–Watson process. For example, Breiman (1968, p. 98) demonstrates the following is true: consider a sequence of nonnegative random variables $X_1, X_2, \ldots$, for which 0 is absorbing in the sense that $X_n = 0$ implies $X_{n+1} = 0$. Assume that there is always a risk of extinction in the following way: For any x, there is a $\delta > 0$ such that

$$\Pr[\text{there exists } n \text{ such that } X_n = 0 \mid X_1, \ldots, X_k] \geq \delta,$$

provided $X_k < x$. Then, with probability 1, either there is an n such that all $X_k = 0$ for $k \geq n$, or $X_k \to \infty$, as $k \to \infty$. So, the process either becomes extinct or grows

indefinitely. We will consider more specific forms of this law for the Galton–Watson process (Theorems 3.2 and 3.3).

The next natural question is, what is the rate of the unlimited growth? It can be answered within the generation counting framework of Galton–Watson type processes (for biological examples, see Sect. 3.2 and following sections). In a more general setup, we must know not only how many children parents get but also the ages at child-bearing (even if they are constant and equal to 1, as in the Galton–Watson process case). In an even more general framework, the iid random variables describing reproduction have to be replaced by iid point processes, and the probabilistic addition of random variable by the superposition of point processes (Sect. C.1). In all these frameworks, in the supercritical case, when the average number of progeny of an individual is greater than 1, the growth pattern is asymptotically exponential. The parameter of this exponential growth is the famous Malthusian parameter. In the supercritical case, we can not only answer questions about the rate of growth but also questions about the asymptotic composition of nonextinct populations. What will the age distribution tend to be? What is the probability of being firstborn? The average number of second cousins? Or the distribution of the times back to an nth grandmother's birth? Very important for biological applications, many of these questions do not have natural counterparts in deterministic models of unlimited growth.

Many other composition questions cannot be posed if we assume that all individuals are of one and the same type. Thus, we are naturally brought on to multitype branching populations: Whenever an individual is born, we know not only its parent's age but also its own type. This latter might be identified, in the most general case, with individual's genotype.

Mathematically, the individual reproduction process then turns into a point process on the product space, type $\times$ age. And the evolution of the newborn's life will no longer be decided in an iid fashion but rather according to a probability kernel, determined by the type of the newborn. The introduction of various types of individuals can be viewed as taking the step from independence to the simplest form of dependence in probability theory, the Markov dependence. One is born of one's parent, who decides when one is to come into this world and also passes on a genotype. Given these two inherited properties, one leads one's life independently of all one's ancestors. This is the Markov model of population growth, the outcome of a straightforward combination of a vague population growth pattern with Markov dependence of random lives and reproduction.

Another general question is what mathematical tools should be used to measure populations. Usually, we are interested in the number of individuals present at a given time. However, we might wish to count only individuals above a certain age. In some applications, we might be interested in the total number of individuals ever born. All these variants are covered under the general concept of additive measures of population size, which goes back almost three decades (Jagers 1975). In this approach, each individual is measured by a random characteristic, a stochastic process, whose value at time t is determined by the individual's type, the individual's age now at time t, and the individual's, and possibly all her progenies' life careers. If the characteristic is assumed to vanish for negative ages, individuals are not taken

into account before they are born. The measure of the population at time t is the sum of all the characteristics evaluated for all the individuals as above. The simplest characteristic is the one that just records whether an individual is born or not, having the value of 1 if it is, and 0 if it is not. It counts all individuals born until time t.

As it will be seen in the sequel, multitype branching processes are a tool for very detailed modeling using the Markov-multitype paradigm. In this setup, the type–space transitions become as important as branching itself. This is very clearly seen in processes with denumerable type spaces, for example, the branching random walk describing gene amplification in Sect. 7.4. The process is a supercritical branching process, but the growth law is not Malthusian: Exponential growth is modified by a negative fractional power factor. Another example is the model of telomere shortening (Sect. 7.7). The transition law is reducible, which produces a variety of unusual behaviors including polynomial growth.

A relatively new application of branching processes in genetics and evolution is the characterization of genealogies. In this approach, a sample of individuals from the process is considered at a given moment t and, conditionally, on this information, the distributions of past events related to the process are sought. The part of the process, existing before t, which contributed to the sample (i.e., excluding the individual whose descendants became extinct before t), is called the reduced process. Examples can be found in Sect. 8.1. The "backward-look" reduced-process limit laws for the critical and subcritical cases are quite different from those in the forward approach.

In classical population genetics models, the population size is a deterministic function of time. The stochastic part of the model is concerned with dependencies between the genetic makeups of the individuals. However, subpopulations of larger populations can be approximated by branching processes (Nagylaki 1990). This has important applications since various rare genotypes, for example, mutant carriers of a rare genetic disease, belong to this category. One such application is gene mapping (Kaplan et al.1995), i.e., determining the location of unknown genes based on their co-inheritance with known (marker) loci.

Branching processes are a conceptually simple tool for modeling diverse aspects of biological populations, not limited to demography, but reaching into cell and molecular biology, genetics, and evolution theory. They provide a framework for detailed considerations, allowing quantitative predictions, beyond metaphorical representations. With a future influx of new detailed biological data, their importance for modeling is likely to increase.

Chapter 2
Biological Background

This chapter is a brief introduction, for mathematicians, to genes, cells, and cancer. It provides general descriptions of the biological phenomenas that are the subject of the mathematical applications developed in later chapters. More specific information relevant to each application is given at the beginning of the section containing the application. No knowledge of biology or chemistry is assumed beyond that learned in high school and forgotten due to disuse. Many biological details are omitted for lucidity. Readers familiar with the biological topics in this introduction may proceed directly to the later chapters.

2.1 Genomes: Changes in DNA and Chromosomes

2.1.1 Genome

The term "genome" refers to the molecules that function in the storage, expression, and inheritance of information in biological systems. The genome of humans and other organisms is dynamic. The number and sequence of its subunits can undergo rapid changes within a few generations.

2.1.2 DNA and Genes

Deoxyribonucleic acid (DNA) is the chemical that is the primary genetic material in the genome of all cells. It is responsible for the storage and inheritance of genetic information. DNA is a polymer consisting of two long complementary strands (Fig. 2.1). Each strand contains a linear sequence of four monomer subunits called bases. The bases are abbreviated *A, T, G,* and *C*. Each *A* base on one strand pairs with a *T* base on the other strand, and each *G* base pairs with a *C* base. The total length of DNA in the genome of each mammalian cell is about 3×10^9 bases. Genes are specific subsequences of DNA that code for proteins. A given gene is typically 10^3–10^4

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_2

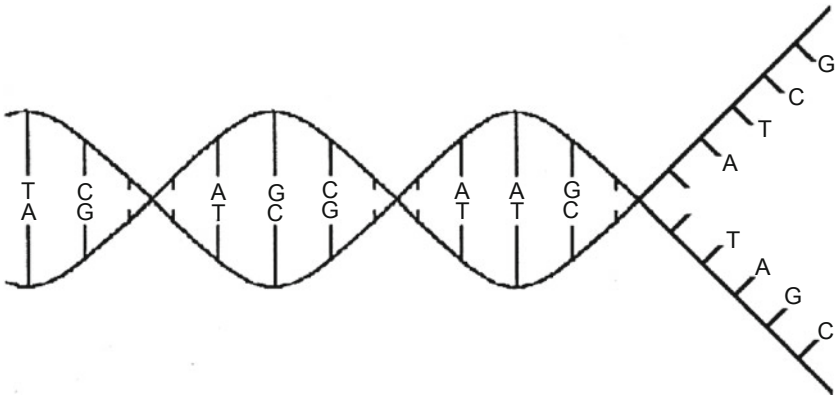


Fig. 2.1 DNA structure. Pairs of complementary bases (*A* and *T*, *G* and *C*) hold together the double-stranded helix of DNA. The two strands are separated on the right, for replication, described further in Fig. 2.3. The sequence of bases in DNA is transcribed into a sequence of bases in RNA and translated into a sequence of bases in protein which functions to determine the observable traits of the organism. Mutational changes in the sequence of bases in DNA result in changes in observable traits which are inherited

bases long and occurs one time or only a few times in the genome. The mammalian genome contains approximately 21,000 different genes that code for proteins. The expression of the genetic information in DNA is accomplished by transcribing a sequence of bases in DNA into a sequence of bases into a related molecule, ribonucleic acid (RNA). The sequence of bases in RNA is then translated into a sequence of amino acids, the subunits of proteins. The proteins carry out catalytic and structural roles which result in the biological properties of cells and organs. So, the flow of information is usually as follows:

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein} \rightarrow \text{phenotype},$$

in words :

$$\text{gene} \rightarrow \text{message} \rightarrow \text{catalyst} \rightarrow \text{observable trait}.$$

2.1.3 Mutation

An alteration in the sequence of bases in DNA is referred to as a mutation. The mutation may be as small as a change in a single base or as large as a rearrangement of most of the DNA in a chromosome. A mutation in DNA may result in an altered sequence of amino acids in protein and/or an altered amount of protein. This may result in a change in the ability of a protein to function properly, resulting in altered properties of cells and organisms. Many mutations are deleterious, but others may

be advantageous or neutral. Examples of altered properties of cells containing mutations include misregulation of cell growth and division leading to malignant tumors, and new capability of mutant cells to grow in the presence of a drug that would kill normal cells. A multitype process model describing mutations as occurring in several possible steps, and an improved method of estimating mutation rates from experimental observations are given in Sect. 6.1.

2.1.4 Noncoding Sequences of DNA

Genes account for a small portion of the genome of mammals. Only about 5–10% of the DNA codes for proteins, the remainder is referred to as the noncoding portion of DNA. The function of the noncoding DNA is only partially understood. Some noncoding regions specify the sequence of bases in RNA that is never translated into protein. Another small portion consists of special DNA sequences that are required to maintain the ends of DNA called telomeres. Maintenance and loss of telomere sequences are discussed as a Bellman–Harris process with denumerable type space in Sect. 7.7. Yet other noncoding sequences, centromeres, are required for the accurate segregation of the DNA into progeny cells. Fragments of DNA that do not contain centromeres distribute into progeny cells in uneven numbers. This is modeled as a single-type Galton–Watson branching process in Sect. 3.7 and as random walk with absorbing boundary in Sect. 7.4.

2.1.5 Repeated Sequences of DNA

Much of the noncoding mammalian DNA consists of sequences which are repeated many times in the genome, most of unknown function. Some repeated sequences are tandemly distributed, others are dispersed throughout the genome. The length (number of bases) of a tandemly repeated unit may be as short as one base or longer than 10^3 bases. The number of repeated units may be as small as two or larger than 10^2. An increase in the number of tandemly repeated units is referred to as amplification, and decrease as deamplification.

The emergence of periodicities of tandemly repeated sequences in DNA by recombination slippage, simulated by a discrete stochastic dynamical system, was discussed by Baggerly and Kimmel (1995). Repeats may also arise by other mechanisms as discussed in Sect. 3.8. See also Bat et al. (1997).

For example, *Alu* elements are DNA sequences of about 300 base pairs. They do not code for proteins. They appear only in mammals. The number of the *Alu*-repeated elements per genome has increased during the evolution of primates. The human genome contains more than one million *Alu* repeat elements dispersed throughout the genome, some with slightly different sequences. Amplification of *Alu*

repeat elements is modeled as a discrete-time Griffith and Pakes branching process, Sect. 7.10.1.

2.1.6 Gene Amplification

Regions of DNA may undergo an increase in number (amplification) or decrease in number (deamplification). The regions of DNA that undergo amplification and deamplification may contain genes, or contain no genes. Amplification and deamplification of regions of DNA-containing genes can result in increases or decreases of amounts of proteins necessary for cell functions. Overproduction of rate-limiting proteins may confer new properties on cells. For example, if the protein is involved in cell proliferation, the cells with an increased amount of this protein may grow as malignant tumors. As another example, if the protein is the target of a toxic drug, then an increased amount of the protein may allow the cells to be resistant and grow in the presence of the drug. Models for gene amplification resulting in tumor cell growth and drug resistance are the subjects of Sects. 3.7, 7.4 and 7.5.

Some inherited human syndromes, such as predisposition to some cancers and neurological diseases, have been related to a rapid change in the numbers of copies of DNA sequences. An unusual aspect of these is the apparently explosive increase in numbers of copies of some sequences from one generation to another. This increase has been modeled as an iterated Galton–Watson process in Sect. 3.8. In contrast to these cases of concerted increases in gene copy numbers, there are situations in which the number of amplified genes is unstable and decreases. Unstable decreases in numbers of amplified genes are modeled as a subcritical Galton–Watson process in Sect. 3.7, and as a branching random walk with absorbing barrier in Sect. 7.4.

2.1.7 Chromosomes

In human cells, the DNA of the genome in divided into 23 pieces of various lengths, each containing large numbers of different genes. Each piece of DNA is folded compactly and associated with proteins and RNA to form a chromosome (Fig. 2.2). In human cells, there are 23 pairs of chromosomes. Each chromosome contains one double-stranded piece of DNA from end to end. The ends of chromosomes are called telomeres. They have special sequences and structures that are necessary for the replication of the end of DNA and the maintenance of chromosomes. DNA in chromosomes replicates and then the products separate from each other in a process called mitosis. Special DNA sequences near the center of chromosomes (centromeres) function to segregate one of each pair of duplicated chromosomes into each of two progeny cells during cell division. This process assures that each progeny cell receives one copy of each chromosome and its associated DNA-containing genes. Fragments of DNA without centromeres may increase in number (replicate) to more

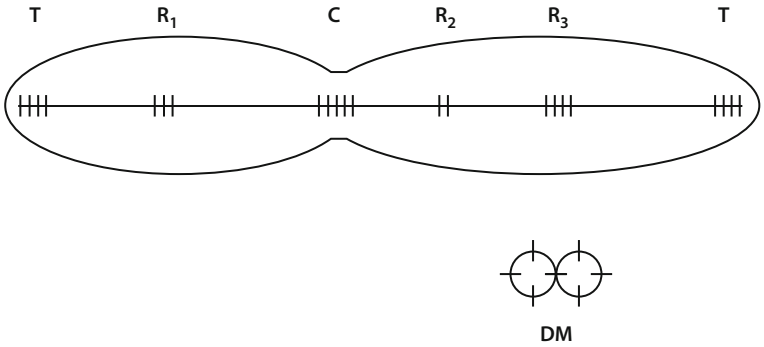


Fig. 2.2 Chromosome. One double-stranded piece of DNA, represented here by a single horizontal line, extends from one end of a chromosome to the other. Several classes of repeated sequences of bases are represented. Telomere (*T*) repeats at each end of the chromosome function to maintain chromosome ends. Centromere (*C*) repeats function to separate chromosomes at mitosis and cell division. Other repeated (*R*) sequences of bases are dispersed throughout the chromosome. Some function to code for proteins (e.g., genes), others are noncoding sequences. Repeated sequences may exist as extrachromosomal elements, also called double minute (*DM*) chromosomes because of their appearance. They may replicate but are not partitioned evenly to progeny cells because they lack the centromeres of chromosomes. The number of repeated units may be variable

than two copies per cell but without centromeres there is no mechanism to distribute exactly equal numbers to each progeny cell.

2.1.8 DNA Replication

DNA replication occurs by a so-called semiconservative mechanism. Two complementary parental strands of DNA separate and each strand forms a template for the production of a new complementary progeny strand. Usually, replication is initiated by the local separation of two strands, the replication fork, and then proceeds along the DNA. The result is two double-stranded DNA molecules, each molecule containing one old strand and one complementary new strand (Fig. 2.3).

Two types of errors in DNA replication have been proposed to result in amplification of repeated DNA sequences, replication slippage, and replication reinitiation. Replication slippage may occur when repeated DNA sequences on one strand fold back on themselves forming a hairpin-like structure. This may cause slippage of the replication complex along one strand relative to the other and resulting in stuttering and repeated replication of a portion of the DNA sequence. The generation of unstable numbers of DNA repeats by replication slippage may contribute to the explosive increase in numbers of repeated sequences in certain cancers and inherited neurological diseases. A mathematical model describing amplification of repeats by replication slippage has been developed (Bat et al. 1997) but is not described as an application here because it is not a branching process. Replication reinitiation

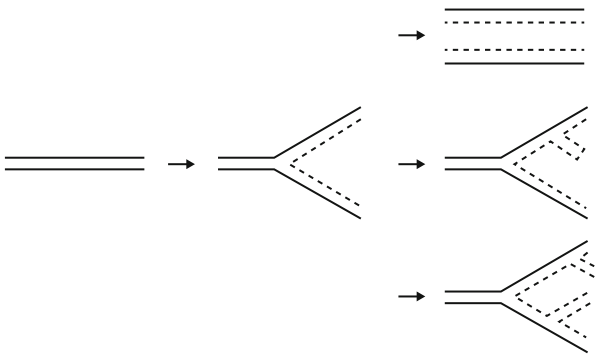


Fig. 2.3 DNA replication. Double-stranded DNA (*left*) replicates by a semiconservative mechanism. The parental strands separate (*center*) and each codes for a new complementary strand. This results in two progeny double-stranded DNA molecules, each containing one old strand and one new strand (*right top*). Two types of errors in DNA replication may result in locally repeated regions (repeat sequences). These errors include slippage and fold back forming hairpin-like structures (*right center*), and replication reinitiation forming branches within branches (*right bottom*)

is another possible mechanism that may contribute to gene amplification. It is visualized as the start of a new replication fork before the previous replication fork has completed moving through the DNA. This leads to the formation of branched DNA structures. Gene amplification by replication reinitiation has been modeled in Sect. 3.8.

2.1.9 Recombination

Recombination is an exchange of pieces of DNA (Fig. 2.4). Recombination can result in new combinations of genes, and an increase or decrease in the numbers of genes. Recombination occurs during the formation of germ cells for sexual reproduction (meiosis) and the division of nonsexual somatic body cells (mitosis). If replicated parts of chromosomes, called chromatids, align properly before recombination, and exchange occurs, then new combinations of genes may occur and be segregated into sex cells. Sometimes parts of chromatids misalign before recombination. Such recombination with misalignment can result in an increase or decrease in the numbers of genes on chromosomes in either meiosis or mitosis . Recombination misalignment leading to gene amplification or deamplification is modeled as a Markov chain with denumerable infinity of states in Axelrod et al. (1994), and simulated as a discrete stochastic dynamical system in Baggerly and Kimmel (1995). Recombination within loops of DNA on the same chromosome may yield small fragments containing genes but not centromeres. When the acentric fragments replicate, amplified numbers of genes may be produced in tandem arrays. If these pieces of DNA recombine and reintegrate into a larger chromosomal piece containing a centromere, then the tandem arrays of amplified genes can become stabilized. This is modeled as a Galton–Watson process with denumerable type space in Sect. 7.5.

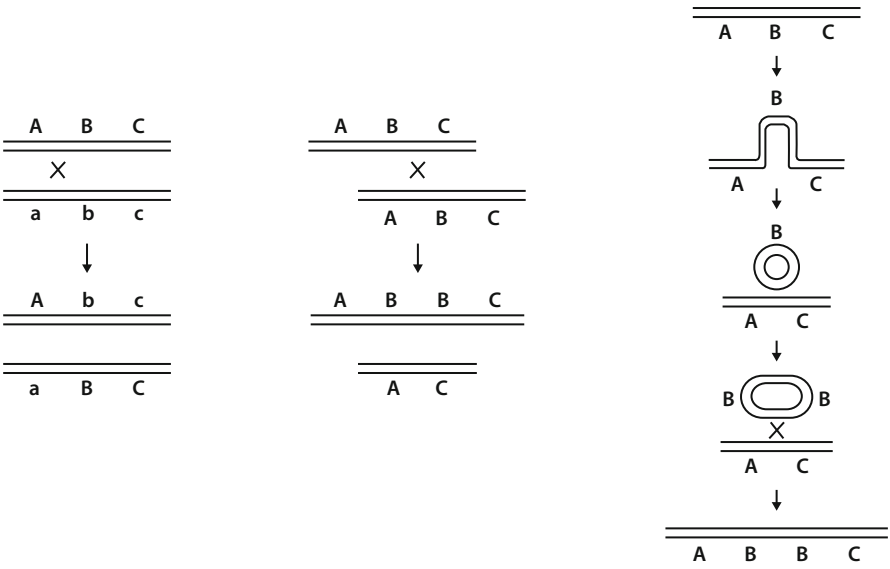


Fig. 2.4 Recombination. New combinations and numbers of genes may be formed by rearrangement of pieces of DNA. Three examples are shown. Double-stranded DNA is represented by *double lines*, genes are represented by *letters*, exchange is represented by an *X*. *Left*, the DNA molecules exchange genes, *upper case* from the mother and *lower case* from the father, to produce new combinations of genes which may then be passed on to progeny. *Center*, the DNA strands slip and misalign before recombination producing one molecule with an increased number of a gene and another molecule with a decreased number of the gene. *Right*, one molecule of DNA undergoes exchange with itself producing a circular piece of DNA. If this piece replicates and then reintegrates, then the result may be an increased number of copies of a gene

2.2 Cells: Cell Cycle Kinetics and Cell Division

2.2.1 Cells as the Basic Units of Life

The basic structural unit of biological function and reproduction is the cell. Mammalian cells are in the range of 20×10^{-6} m of size although there are many cells of different functions and different shapes that are smaller or larger. The structure of the cell is a series of bag-like compartments with specialized functions. The "bags" are made up of membranes that function as barriers, and for transport of molecules. The innermost compartment is the nucleus which contains highly compacted DNA and accessory molecules for expression of genes. Outside of the nucleus is a series of compartments for the synthesis and degradation of molecules used for catalysis, structure, and energy generation. The outermost cell membrane, and its accessory molecules, also function as barriers and for transport, and in addition, for communication with other cells. Communication between cells can occur via small molecules that diffuse between cells such as hormones, or via molecules that become fixed to the surface of other cells, such as antigens which function in the immune system. A model for multivalent antigen binding as a multitype Galton–Watson process is given in Sect. 6.5.

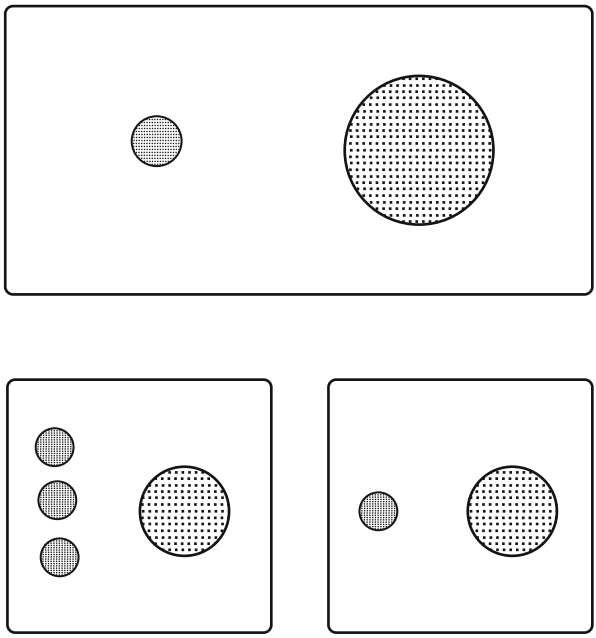


Fig. 2.5 Cell division and partition of contents. During the growth of cells the amount of DNA in the nucleus (*large dark circle*) doubles and is partitioned evenly *into two daughter cells* at cell division. However, other cell constituents may not exactly double and may not partition evenly, resulting in cells with different numbers of these constituents. These constituents include extrachromosomal pieces of DNA, subcellular organelles, and intracellular parasites

2.2.2 Cell Growth, Division, and Death

In a multicellular organisms, adult tissue homeostasis is maintained. The number of new cells produced by cell division is balanced by the the number of differentiated cells that die and are removed. For instance, in the colon, some differentiated cells are constantly being removed and excreted with the feces. Other cells die by a conservative programed cell death called apoptosis, in which pieces of the cells are engulfed by specialized neighboring cells. In the adult nervous system, dividing cells can give rise to several cell types, and a subset of newly born cells are culled, undergoing cell death via apoptosis. The production of several cell types and the death of some have been modeled as a multitype branching process in Sect. 6.2.

Cells grow in size and divide into two. The DNA in the nucleus exactly doubles in amount, is packaged into chromosomes, and is then partitioned evenly between two progeny cells at cell division. However, other processes are less exact. The size to which cells grow before they divide is not exactly the same for all cells of a given type, the lifetimes of cells at division are not exactly the same for all cells, and the non-DNA materials are not partitioned exactly between the two progeny cells (Figs. 2.5 and 2.6). The distribution of cell sizes and cell lifetimes may be stable over time for

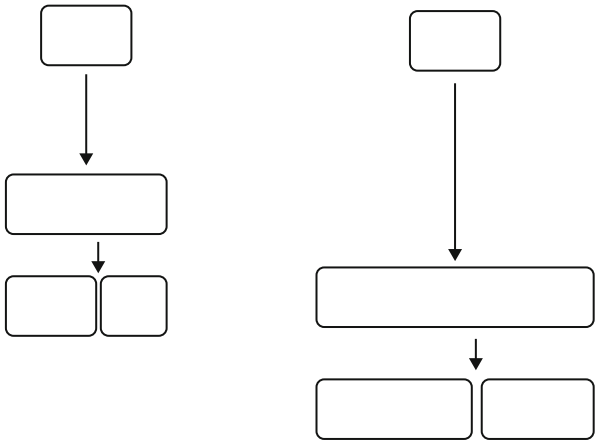


Fig. 2.6 Cell division and cell size. Cells may grow for different times and obtain different sizes before they divide. At cell division, cells may divide asymmetrically resulting in progeny sibling cells of different size

a population of one cell type, but differ for populations of cells of different types. Apparently, mechanisms exist to maintain these parameters within a population of cells of one type. Populations of cells with different values of parameters may differ in important characteristics, such as whether or not they are malignant. A Galton–Watson model describing the growth and division, and death and quiescence of cells is given in Sect. 3.2. Another model in the form of a Galton–Watson process with continuous type space is described in Sect. 7.8.1.

During development of multicellular organisms some cells divide into two cells which differ in shape and function. This situation is modeled as multitype branching process in Sects. 6.3 and 7.8.2. If fragments of DNA are not connected to chromosomes they may not exactly double in number before each cell division and may not partition exactly into the two progeny cells. Entities such as subcellular organelles or intracellular parasites can divide within dividing cells. An appropriate model for this is a Markov process model of infinitely many types. Such a model exhibits quasistationarity, as discussed in Sect. 7.6.

2.2.3 Stem Cells

Multicellular organisms are composed of many specialized cells that differ in function. Totipotent stem cells in the early embryo can yield all of the different specialized cells in a multicellular organism. Multipotent stem cells found in the adult can yield only a limited number of different kinds of cells. An example of the latter are the hematopoietic stem cells found in the bone marrow that yield the many specialized cells in the blood. The purpose of the bone marrow transplant procedure is to replenish active hematopoietic stem cells that are capable of yielding all of the

different specialized cells in the blood. Stem cells may be quiescent and may not divide, or if stimulated, may become active and divide. Active stem cells may divide symmetrically to produce two stem cells, or divide symmetrically to produce two differentiated cells, or divide asymmetrically to produce one stem cell and one differentiated cell. Stem cell division has been modeled as age-dependent and multitype branching processes in Sect. 7.8.1.

2.2.4 Cell Cycle Kinetics

The time period between cell birth and cell division is referred to as the cell cycle time. Several distinct events or phases can be distinguished during each cell's lifetime. The first event is the birth of two progeny cells at cell division, also called cytokinesis or mitosis, abbreviated M. The time gap between the birth of a new cell and the initiation of DNA synthesis is called gap one, abbreviated G_1. The period of DNA synthesis is abbreviated S. The time gap between S and the next mitosis is abbreviated G_2. After G_2, during the next M phase the cell divides to form two new cells. The sequence of phases M, G_1, S, G_2, and M repeats in progeny cells of each subsequent generation, and thus the name cell cycle. For mammalian cells, a typical cell cycle time may be 12–24 h, or even longer. For a cell cycle time of 24 h, the duration for the cell cycle phases M, G_1, S, and G_2 might be 0.5, 8, 12, and 3.5 h. The duration of the G_1 phase is usually the most variable portion of the cell cycle. Cells which have longer cell cycle times, either because of genetics, environment or developmental fate, usually have equally extended G_1 time periods, although important exceptions exist. The relative duration of the cell cycle phases in a growing population of cells can be inferred from the percentage of cells with different amounts of DNA, or from the rate at which cells accumulate in one phase of the cell cycle when blocked with a phase-specific drug (stathmokinesis). Cell cycle kinetics are modeled as a Bellman–Harris process in Sect. 5.4, and as a Markov time-continuous branching process in Sect. 4.2.

2.3 Cancer

2.3.1 Cancer Cell Populations Are Immortal

Cancer is a problem in persistent cell proliferation. Tumors are populations of cells that accumulate in abnormal numbers. The increased number of cells is due to an increased ratio of cell birth rate over cell death rate. Cancer cells do not necessarily grow faster than normal cells, but they are persistent. They may not stop dividing under conditions where normal cells would stop, and/or they may not die under conditions where normal cells would die. Normal cells in an adult seem to be capable of a finite number of divisions and then the lineage dies out. The process of a cell lineage losing proliferative potential is referred to as senescence. Some cells in

a tumor are capable of an indefinite number of divisions so that the lineage can persist. Populations of cells that divide without limit are referred to as immortal. The mechanisms controlling senescence and immortality are partially known. For instance, an inhibitor has been identified in old senescent cells that is not expressed efficiently in young cells. In addition, there seems to be a difference between many normal cells, which are capable of senescence, and tumor cells which are immortal, viz., a difference in the ability to maintain the ends of chromosomes. The ends of chromosomes contain repeated sequences of DNA called telomeres. Although most of the length of DNA is duplicated exactly once during each cell cycle, that is not always true of the repeated DNA sequences in telomeres at the ends of chromosomes. The telomeric DNA sequences can increase or decrease in length at each round of DNA replication. Normal cells seem to progressively lose the telomeric repeat sequences and senesce (age), whereas some tumor cells seem to maintain them and continue to divide. This has been modeled as a Bellman–Harris process with denumerable type space in Sect. 7.7.

2.3.2 Tumor Heterogeneity and Instability

Tumors are derived from single cells. This conclusion has resulted from observations in which all the cells in a tumor share a common change from normal cells. Cells from different tumors have different changes. The changes observed range in size from single base mutations in DNA to large chromosome rearrangements. In addition to the common changes among the cells in a single tumor, many cells may show additional changes distinct for each cell in a tumor. In other words, tumors are monoclonal in origin but heterogeneous. Many tumors are genetically unstable showing an increased probability of undergoing mutations or gene amplification. A mutant gene may produce a protein product with an altered function, and a gene with amplified number of copies may produce an increased amount of a protein. If the protein is the target of a toxic drug then a tumor cell producing an increased amount of this protein may become resistant to this drug and escape effective chemotherapy. Gene amplification leading to drug resistance has been modeled using the Galton–Watson process in Sect. 3.7, modeled as a Galton–Watson process with denumerable types in Sect. 7.5, and as a branching random walk with absorbing barrier in Sect. 7.4.

2.3.3 Cell Cycle and Resistance to Chemotherapy

Some forms of cancer chemotherapy attempt to exploit differences between the cell cycle of normal cells and malignant cells. For instance, a single drug that inhibits DNA synthesis would be expected to kill more tumor cells than normal cells, if more tumor cells than normal cells are synthesizing DNA during the period of chemotherapy. Sometimes, two or more drugs are used which affect different cell cycle phases, or have different mechanisms of inhibition (combination chemotherapy). The purpose is to overcome possible resistance to a single drug and to increase the probability

of catching tumor cells in different phases of the cell cycle. Therefore, it is important to be able to determine the cell cycle phase durations of normal and malignant cells, the cell cycle phase specificity of drugs, and the changes that occur when cells are exposed to anticancer drugs. A multitype process is used to model changes in the cell cycle during chemotherapy in Sect. 6.4. The emergence of cross resistance (each cell resistant to two drugs) is modeled as a time-continuous branching process in Sect. 4.2.

2.3.4 Mutations in Cancer Cells

Rates of mutations that occur in cancer cells are estimated by a method called the fluctuation test. The procedure was originally developed for bacteria. In this test, the following are observed: the number of mutant cells arising in many parallel cultures, the number of cell divisions in the cultures, and the number of cultures which contain no mutant cells. The original method of calculating mutation rates from these observations is based on the assumption that each mutant cell resulted from a single rare event that is irreversible. This was appropriate for the bacterial mutations originally observed, but not for many mutations in cancer cells. The mutations may not be rare, may not be irreversible, and may not be due to a single step. Application of the fluctuation test to cancer cells required the development of methods that took into account these possibilities. Multitype branching processes were used to model two-step mutations and interpret data from the fluctuation test, see Sect. 6.1.

2.3.5 Tumor Progression

Clinically detected tumors are composed of a million or more cells that may be a colony derived from one cell. Evidence for the monoclonal origin of tumors is based, in part, on the observations of chromosome aberrations in chronic myeloid leukemia, and DNA polymorphisms in colon cancer. Each cell in the tumor population has the same change from normal cells, suggesting that the change occurred in the founding cell. This founding cell is sometimes referred to as the cancer stem cell. (The term cancer stem cell has also been used to refer to the cells in a tumor that continue to divide and expand the tumor.) Although the tumor cell population may be clonal, by the time the tumor is detected, the population of cells may be heterogeneous. Tumor heterogeneity refers to the observation that cells within a tumor may differ from each other for one or more characteristics such as DNA sequences, expression of subsets of genes, expression of proteins, and resistance to chemotherapeutic drugs. The evolution of tumor heterogeneity is referred to as tumor progression. It results from the accumulation of mutations (alterations in the sequence of DNA bases) and epigenetic changes (modifications to DNA bases and associated proteins) as progeny of the founding cell divide many times. Mutations to drug resistance have been modeled as two-type branching process in Sect. 5.5.2.

2.4 Population Genetics and Evolution

2.4.1 Wright–Fisher Model and Coalescent-Based Models

Two well-known models of population genetics are the Wright–Fisher model and the coalescent-based model (Barton 2007, Chap. 15). The Wright–Fisher model for genetic drift assumes that the population size is constant, the generations are nonoverlapping, there is random mating between individuals with different alleles neutral to selection, random sampling, and no additional mutant alleles occur as the population evolves. The Wright–Fisher model views changes among individuals as a population evolves forward from an initial population. In contrast, the coalescent-based model views a population backward. A sample of differing individuals in the final population is analyzed, and this information, assuming no selection, is used to infer their most recent common ancestor (MRCA) and the time (in generations) to the MRCA (T_{MRCA}). These models are discussed in Chap. 8.

2.4.2 Human Immunodeficiency Virus

Acquired immune deficiency syndrome (AIDS) emerged in the early 1980s, and has since been associated with more than 25 million deaths worldwide. The causative agent of AIDS is the human immunodeficiency virus (HIV). It is an RNA retrovirus that infects cytotoxic T-cells, an essential component of the immune system. The virus may multiply in the T-cells and inhibit their function. Or, the virus may produce a DNA copy that integrates into the DNA of the host T-cell and become latent, reproducing with the DNA of the host T-cell when the host T-cell divides. The latent DNA copy of the virus may, at a later time, become active and produce an RNA copy, which in turn can produce new viruses to infect other cells. The polymerase enzymes that copy RNA to RNA (RNA-dependent RNA polymerase), or that copy RNA to DNA (reverse transcriptase) are extremely error prone, in the range of 0.1–0.3 mutations per genome per replication. This high mutation rate and natural selection have resulted in the well-documented diversity of genome sequences, both within individuals and between individuals in different geographical locations. Some of these differences have been described as different subspecies or quasispecies, and have been partitioned into discrete clades on phylogenetic trees (Holmes 2009). Some of the mutations (escape mutations) allow virus mutants to resist immunotherapy or to resist single antiviral drugs. The genealogies of mutants resistant to immunotherapy or to drugs, and the brief appearance of low levels of viruses in the blood of treated patients, have been modeled as multitype Galton–Watson processes and are discussed in Chap. 6.9.5.

2.5 References

References cited here include textbooks and monographs on the topics discussed in this chapter. Specific citations to the primary literature are given in the applications sections of later chapters.

2.5.1 Textbooks and Monographs in Biology

Two outstanding textbooks in molecular and cellular biology used in upper level undergraduate and graduate classes:

- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2008) Molecular biology of the cell, 5th edn. Garland, New York
- Lodish H, Berk PA, Kaiser CA, Krieger M, Scott MP, Bretscher A, Ploegh H, Matsudaira P (2008) Molecular cell biology, 6th edn. Freeman, New York. Ch. 25 Cancer, pp 1107–1150. (Contents of a previous edition are searchable online at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books&itool=toolbar)

Additional information on genomics, DNA, and chromosomes:

- Brown TA (2002) Genomes, 2nd edn. Wiley-Liss, Oxford (Entire contents available online at http://www.ncbi.nlm.nih.gov/books/NBK21128/)
- Goldstein DB (ed) (1999) Microsatellites: evolution and applications. Oxford University Press, New York
- Kornberg A, Backer TA (2005) DNA replication. University Science Books, Herndon
- Krebs JE, Goldstein ES, Kilpatrick ST (2014) Lewin's genes X. Jones & Bartlett, Sudbury
- Lesk AM (2011) Introduction to genomics, 2nd edn. Oxford University Press, New York
- Lynch M (2007) The origins of genome architecture. Sinauer, Sunderland
- Watson JD, Baker T, Bell S, Gann A, Levine M, Losick R (2008) Molecular biology of the gene, 6th edn. Pearson Benjamin Cummings, Menlo Park
- Zhang W, Shmulevich I (2006) Computational and statistical approaches to genomics, 2nd edn. Springer, New York

Informative textbooks on cancer:

- DeVita VT, Lawrence TS, Rosenberger SA (eds) (2010) Cancer: principles & practice of oncology. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia
- Komarova NL, Wodarz D (2013) Targeted cancer treatment in silico: small molecule inhibitors and oncolytic viruses. Birkhäuser

- Kufe DW, Pollock RE, Weichselbaum RR, Bast RC Jr, Gansler TS, Holland JF, Emil E (eds) (2003) Cancer medicine, 6th edn. Decker, New York (Table of contents available online. Entire contents searchable online at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books&itool=toolbar)
- Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson CB (eds) (2008) The molecular basis of cancer, 3rd edn. Saunders, 704 pp
- Pelengaris S, Khan M (2013) The molecular biology of cancer: a bridge from bench to bedside. 2nd edn. Wiley-Blackwell, Chichester
- Weinberg RA (2014) The biology of cancer, 2nd edn. Garland Science, New York
- Wodarz D, Komarova NL (2005) Computational biology of cancer: lecture notes and mathematical modeling. World Science

Journal articles reviewing cancer:

- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646–674
- Nature Reviews: Cancer (many issues, search by subject)

An excellent textbook on human molecular genetics:

- Strachen T, Read A (2010) Human molecular genetics, 4rd edn. Garland Science, Independence (Third edition, 2003, contents searchable online at http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.TOC&depth=2)

Textbooks that describe experimental techniques in molecular genetics with informative visuals:

- Micklos DA, Freyer GA, Crotty DA (2003) DNA science: a first course. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Brooker R (2011) Genetic analysis and principles. McGraw-Hill, New York

Websites focusing on cancer:

- National Cancer Institute's Cancer Information (types of cancer by body location/system, alphabetical list of cancers) http://cancer.gov/cancerinformation
- National Cancer Institute (many resources for researchers, cancer information, clinical trials, statistics, research programs, research funding, about NCI) http://www.cancer.gov

Websites with searchable databases of biological and medical information, and databases with searchable biological and medical published literature:

- National Center for Biotechnology Information (PubMed, Genome Wide Association Studies, Online Inheritance in Man, other databases, and entire contents of books). http://www.ncbi.nlm.nih.gov/
- Medline database of biomedical literature (Follow links: Advanced, Field [Title/Abstract], search terms subject heading "models, theoretical", or search key word "mathematical model" or "branching process"). http://www.ncbi.nlm.nih.gov/pubmed/

2.5.2 *Mathematical Biology*

Several textbooks provide examples of mathematical approaches to biological problems, although most examples are deterministic rather than stochastic:

- Adler FR (1998) Modeling the dynamics of life: calculus and probability for life scientists, 2nd edn. Brooks/Cole, Belmont (Chap. 6.2 Stochastic models. Chap. 7 Probabilistic models)
- Allen LJS (2007) Introduction to mathematical biology. Pearson Prentice Hall, Upper Saddle River
- Batschelet E (1992) Introduction to mathematics for life sciences, 3 rd edn. Springer-Verlag, New York
- Britton NF (2003) Essential mathematical biology. Springer, New York
- Brown D, Rothery P (1993) Models in biology: mathematics, statistics and computing. Wiley, New York
- Eisen M (1988) Mathematical methods and models in the biological sciences: linear and one-dimensional theory. Prenctice Hall, Englewood Cliffs
- Hoppensteadt FC, Peskin CS (2001) Modeling and simulation in medicine and the life sciences, 2nd edn. Springer-Verlag, New York
- Logan JD, Wolesensky WR (2009) Mathematical methods in biology. Wiley, Hoboken (Chap. 7, Stochastic processes)
- Marotto FR (2006) Introduction to mathematical modeling using discrete dynamical systems. Thomson Brooks/Cole, Belmont
- Wakeley J (2009) Coalescent theory: an introduction. Roberts, Greenwood Village (Chap. 2 Probability theory)

Monographs on mathematical biology include the following:

- Lange K (2003) Applied probability. Springer-Verlag, New York
- Durrett R, Durrett R (2010) Probability: theory and examples, 4th edn. Cambridge University Press, Cambridge
- Ewens WJ, Grant GR (2005) Statistical methods in bioinformatics, 2nd edn. Springer, New York
- Levin SA (ed) (1994) Frontiers in mathematical biology. Lecture notes in biomathematics, vol 100. Springer-Verlag, Berlin
- Murray JD (2002) Mathematical biology: I An introduction, 3rd edn. Springer-Verlag, New York
- Murray JD (2002) Mathematical biology: II Spatial models and biomedical applications. 3rd edn. Springer-Verlag, New York
- Segel LA (1987) Modeling dynamic phenomena in molecular and cellular biology. Cambridge University Press, Cambridge
- Segel LA (ed) (1980) Mathematical models in molecular and cellular biology. Cambridge University Press, London.

Introductions to stochastic processes:

- Allen LJS (2011) An introduction to stochastic processes with applications to biology, 2nd edn. CRC Press, Boca Raton
- Lawler GF (2006) Introduction to stochastic processes, 2nd edn. Chapman and Hall, New York
- Karlin S, Taylor HM (1984) An introduction to stochastic modeling, 3rd edn. Academic, San Diego
- Resnick SI (1992) Adventures in stochastic processes. Birkhauser, Boston
- Ross SM (1996) Stochastic processes, 2nd edn. Wiley, New York
- Taylor HM, Karlin S (1998) An introduction to stochastic modeling, 3rd edn. Academic, Harcourt Brace, Boston

More advanced treatments of branching processes:

- González M, del Puerto IM, Martínez R, Molina M, Mota M, Ramos A (eds) (2010) Workshop on branching processes and their applications. Springer, Berlin
- Haccou P, Jagers P, Vatutin V (eds) (2005) Branching processes: variation, growth and extinction of populations. Cambridge University Press, Cambridge
- Harris TE (1963) The theory of branching processes. Springer-Verlag, Berlin
- Jaggers P (1975) Branching processes with biological applications. Wiley, London
- Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic, New York
- Karlin S, Taylor HM (1975) A second course in stochastic processes, 2nd edn. Academic, New York

A survey of some mathematical models of tumor cell growth, chemotherapy, and drug resistance that is useful for mathematicians and readable for biologists:

- Wheldon TE (1988) Mathematical models in cancer research. Adam Hilger, Bristol

Monographs and textbooks on population genetics and evolution:

- Barton NH, Briggs DEG, Eisen JA, Goldstein DB (2007) Evolution. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Brown SM, Hay JG, Oster H (2008) Essentials of medical genomics. Wiley-Blackwell, Hoboken
- Cavalii-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Ewens W (2004) Mathematical population genetics: I. Theoretical introduction, 2nd edn. Springer, New York
- Gillespie JH (2004) Population genetics: a concise guide, 2nd edn. Johns Hopkins University Press, Baltimore
- Graur D, Li W-H (2000) Fundamentals of molecular evolution, 2nd edn. University of Chicago Press, Chicago
- Hamilton M (2009) Population genetics. Wiley, Hoboken
- Hartl DL, Clark AG (2007) Principles of population genetics, 4th edn. Sinauer, Sunderland

- Hein J, Schierup MH, Wiuf C (2005) Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, Cambridge
- Relethford JH (2012) Human population genetics. Wiley, Hoboken
- Takahata N, Clark AG (1993) Mechanisms of molecular evolution. Tokyo and Sinauer Associates, Sunderland
- Wakeley J (2009) Coalescent theory: an introduction. Roberts, Greenwood Village

2.5.3 Arguments for Mathematical Modeling Biological Phenomena, with Examples

Several essays have emphasized the importance of mathematical modeling for progress in molecular and cellular biology:

- Byrne HM (2010) Dissecting cancer through mathematics: from the cell to the animal model. Nat Rev Cancer 10:221–230
- Gatenby RA, Maini PK (2003) Mathematical oncology: cancer summed up. Nature 421:321
- Mogilner A, Wollman R, Marshall WF (2006) Quantitative modeling in cell biology: what is it good for? Dev Cell 11:279–287
- Mukherji S, van Oudenaarden A (2009) Synthetic biology: understanding biological design from synthetic circuits. Nat Rev Genet 10:859–871
- Oates AC, Gorfinkiel N, González-Gaitán M, Heisenberg C-P (2009) Quantitative approaches in developmental biology. Nat Rev Genet 10:517–530
- Reeves GT, Fraser SE (2009) Biological systems from an engineer's point of view. PLoS 7: e21
- Sontag ED (2005) Molecular systems biology & control. Eur J Control 11:396–435
- Tomlin CJ, Axelrod JD (2007) Biology by numbers: mathematical modelling in developmental biology. Nat Rev Genet 8:331–340
- Wilkinson DJ (2009) Stochastic modeling for quantitative description of heterogeneous biological systems. Nat Rev Genet 10:122–133

Chapter 3
The Galton–Watson Process

The Galton–Watson (GW) process is the oldest, simplest, and best-known branching process. It can be described as follows.

A single ancestor particle lives for exactly one unit of time and at the moment of death produces a random number of progeny according to a prescribed probability distribution. Each of the first-generation progeny behaves, independently of each other, as the initial particle did. It lives for a unit of time and produces a random number of progeny. Each of the second-generation progeny behaves in the identical way, etc. From the fact that the lifespans of all particles are identical and equal to 1, it follows that the process can be mathematically described using a discrete time index, identical to the number of successive generation. The particle counts Z_n in the successive generations $n = 0, 1, 2, \ldots$ (where generation 0 is composed of the single initial particle) form a sequence of random variables with many interesting properties, for example, the *Markov property*). Properties of the GW process provide intuitions about more complicated branching.

The simplicity of the GW process makes it an appropriate and frequently employed tool for introductory study of the processes of proliferation in biology. It is applicable whenever the hypothesis of discrete nonoverlapping generations is justified. An example of the GW branching process is the process describing the polymerase chain reaction in Sect. 1.2.

3.1 Construction, Functional Equation, and Elementary Properties

The material in this section follows the style of Athreya and Ney (2004). Let us suppose that the number of progeny produced by each particle is a nonnegative integer random variable with distribution function $\{p_k;\ k = 0, 1, 2, \ldots\}$.

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_3

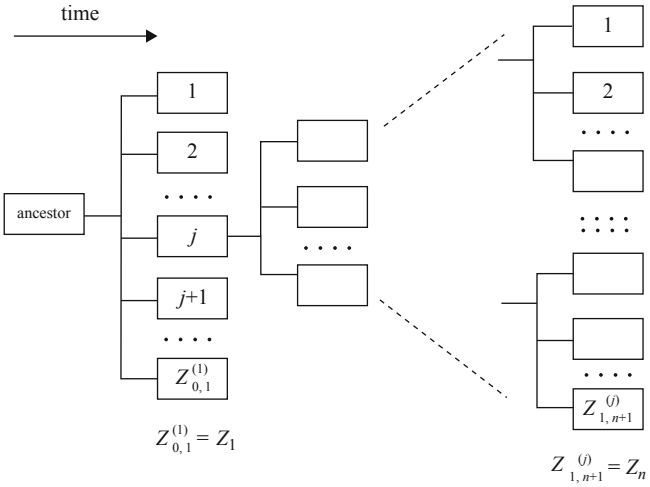


Fig. 3.1 The backward equation for the Galton–Watson process

3.1.1 Backward Equation

Any particle existing in the process, except for the ancestor of the process, can be assigned to a subprocess traceable to a particular first-generation offspring of the ancestor. In other words, the process can be represented as a union of the subprocesses initiated by the first-generation offspring of the ancestor particle.

The number Z_{n+1} of particles in the generation $n+1$ of the process is equal to the sum of the particle counts in the generation n of all the Z_1 subprocesses initiated by the first-generation offspring of the ancestor particle. Let $Z^{(j)}_{1,n+1}$ denote the number of individuals at time $n+1$ in the process started by a single ancestor born at time 1. The additional superscript (j) denotes the jth independent identically distributed (iid) copy. Mathematically, the random variable Z_{n+1} is equal to the sum of Z_1 random variables $Z^{(j)}_{1,n+1}$, or (see Fig. 3.1):

$$Z_{n+1} = \begin{cases} Z^{(1)}_{1,n+1} + \cdots + Z^{(Z_1)}_{1,n+1}, & Z_1 > 0, \\ 0, & Z_1 = 0, \end{cases}$$

or

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z^{(j)}_{1,n+1}. \tag{3.1}$$

Random variables $Z^{(j)}_{1,n+1}$ are (iid copies and their common distribution is identical to that of Z_n. Equation (3.1) can be equivalently written as

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z^{(j)}_n.$$

By the pgf Theorem 3.1 (part 6), it yields the following pgf identity:

$$f_{n+1}(s) = f_1[f_n(s)] = f[f_n(s)]. \tag{3.2}$$

If we note that $Z_0 = 1$ implies $f_0(s) = s$, this yields the following:

$$f_n(s) = f^{(n)}(s) = \underbrace{f\{\cdots f[f}_{n \text{ times}}(s)]\cdots\}, \tag{3.3}$$

i.e., the pgf of Z_n is the nth functional iterate of the progeny pgf $f(s)$.

3.1.2 *Forward Equation*

An alternative approach is based on the fact that any particle in the $(n+1)$-st generation of the process can be traced to its parent in the nth generation of the process. Let X_{in} denote the number of progeny of the ith particle existing in generation n. More generally, let $\{X_{in}\}_{i\geq 1, n\geq 0}$ be a doubly infinite array of iid rv such that $E(X_{10}) = \mathrm{m} < \infty$. Then

$$\begin{aligned} Z_0 &= 1, \\ Z_{n+1} &= \begin{cases} X_{1n} + \cdots + X_{Z_n,n}, & \text{if } Z_n > 0, \\ 0, & \text{if } Z_n = 0, \end{cases} \quad n \geq 1, \end{aligned} \tag{3.4}$$

or

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{in},$$

i.e., the number of individuals (particles, cells), in the nth generation of the process is equal to the number of progeny of all individuals in the generation $n-1$. In the terms of pgf's we obtain a new recursion:

$$f_{n+1}(s) = f_n[f_1(s)] = f_n[f(s)]. \tag{3.5}$$

In the case of the GW process, the above recurrence also leads to Eq. (3.3). However, for more general processes, the forward construction may not be feasible. We will return to this matter.

Nontriviality

We exclude situations in which the number of particles is deterministic or when it is either 0 or 1. Therefore, we assume throughout that $p_0 + p_1 < 1$ and that $p_j \neq 1$ for any j.

3.1.3 Moments

The moments of the process, when they exist, can be expressed in the terms of the derivatives of $f(s)$ at $s = 1$. For the mean we have

$$E(Z_1) = f'(1-) \equiv m,$$

where m is the mean number of progeny of a particle. From the chain rule

$$E(Z_n) = f_n'(1-) = f_{n-1}'(1-)f'(1-) = \cdots = m^n. \tag{3.6}$$

Similarly, using the chain rule for the second derivative, one concludes that

$$\mathrm{Var}(Z_n) = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{m-1}, & m \neq 1, \\ n\sigma^2, & m = 1, \end{cases} \tag{3.7}$$

where $\sigma^2 = \mathrm{Var}(Z_1)$ is the variance of the progeny count. Higher moments are derived similarly, if it exists. The linear growth of variance in the critical case ($m = 1$) is consistent with the "heavy tails" of the distribution of Z_n in the critical case, mentioned in Sect. 1.5.3.

3.1.4 The Linear-Fractional Case

Usually, after several iterations, the functional form of the iterates $f_n(s)$ becomes intractable. The linear-fractional case is the only nontrivial example for which they have been explicitly computed. Suppose

$$p_0 = \frac{1-b-p}{1-p},\ p_k = bp^{k-1},\ k = 1, 2, \ldots .,\ p \in (0, 1).$$

Then

$$f(s) = 1 - \frac{b}{1-p} + \frac{bs}{1-ps}, \tag{3.8}$$

and $m = \frac{b}{(1-p)^2}$. The equation $f(s) = s$ has roots q and 1. If $m > 1$, then $q < 1$, if $m = 1$, then $q = 1$, if $m < 1$, then $q > 1$. The following expressions are proved by induction (for a direct derivation, c.f. Athreya and Ney 2004)

$$f_n(s) = 1 - m^n\left(\frac{1-q}{m^n-q}\right) + \frac{m^n\left(\frac{1-q}{m^n-q}\right)^2 s}{1-\left(\frac{m^n-1}{m^n-q}\right)s},\quad m \neq 1, \tag{3.9}$$

$$f_n(s) = \frac{np-(np+p-1)s}{1-p+np-nps},\quad m = 1. \tag{3.10}$$

The linear functional pgf corresponds to the geometric distribution with a rescaled term at zero.

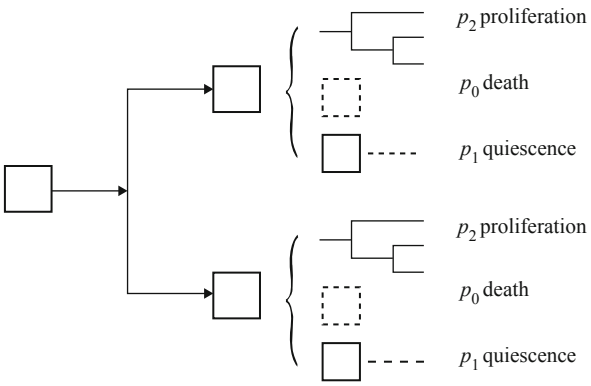


Fig. 3.2 A schematic representation of the cell cycle model. Each of the daughter cells, independently, starts growing with probability p_2, dies with probability p_0, or becomes quiescent with probability p_1. Where $p_0 + p_1 + p_2 = 1$

3.2 Application: Cell Cycle Model with Death and Quiescence

The material of this section follows Kimmel and Axelrod (1991). The fundamental step in the proliferation of a population of cells is the division of one cell into two cells. After completing its life cycle each cell approximately doubles in size and then divides into two progeny cells of approximately equal sizes. Populations derived from single cells are referred to as clones or colonies. It has been experimentally observed that similar cells may not yield colonies with the same number of cells after the same time. This may be due to various factors like the randomness of cell death and quiescence.

3.2.1 The Mathematical Model

We consider a process more general than the standard GW process. It is initiated by a single proliferating cell (Fig. 3.2). This cell divides, and each of its progeny, independently, may (i) become proliferative with probability (wp) p_2, (ii) become quiescent wp p_1, or (iii) die wp p_0. Quiescent cells continue to exist without proliferating nor dying. They may return to active growth and proliferation even after a very long time, or eventually die; in the present model, these possibilities are not considered. We assume $p_0 + p_1 + p_2 = 1$.

The equations describing the model will be recurrences for the probability generating functions (pgfs) of the number of proliferating and quiescent cells present in the population in successive generation, analogous to the backward Eq. (3.1). Let us denote by Z_n the number of proliferating cells in the nth generation and by Q_n the number of quiescent cells in the nth generation. Also, let $Z_{n,k}^{(j)}$ and $Q_{n,k}^{(j)}$ denote, respectively, the number of kth generation proliferating and quiescent offspring of the jth of the Z_n proliferating particles of the nth generation. The number of offspring of

Table 3.1 Derivation of the backward equation for the cell cycle model

(Z_1, Q_1)	Probability	(Z_{n+1}, Q_{n+1})	$f_{n+1}(s, w)$
$(0, 0)$	p_0^2	$(0, 0)$	1
$(0, 1)$	$2p_0p_1$	$(0, 1)$	w
$(0, 2)$	p_2^2	$(0, 2)$	w^2
$(1, 0)$	$2p_2p_0$	$(Z_{1,n+1}^{(1)}, Q_{1,n+1}^{(1)})$	$f_n(s, w)$
$(1, 1)$	$2p_2p_1$	$(Z_{1,n+1}^{(1)}, Q_{1,n+1}^{(1)} + 1)$	$f_n(s, w)w$
$(2, 0)$	p_2^2	$(Z_{1,n+1}^{(1)} + Z_{1,n+1}^{(2)}, Q_{1,n+1}^{(1)} + Q_{1,n+1}^{(2)})$	$f_n(s, w)^2$

a quiescent cell is always equal to 1, the same quiescent cell. We write the following equations:

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_{1,n+1}^{(j)}. \tag{3.11}$$

$$Q_{n+1} = \sum_{j=1}^{Z_1} Q_{1,n+1}^{(j)} + Q_1. \tag{3.12}$$

Let us denote by $f_n(s, w)$, the joint pgf of random variables Z_n and Q_n (see Definition A.1 in the Appendix A). Let us note that $Z_{1,n+1}^{(j)}$ and $Q_{1,n+1}^{(j)}$ have distributions identical to those of Z_n and Q_n, respectively. To obtain the recurrence for the pgf, we first condition (Z_{n+1}, Q_{n+1}) on given values of (Z_1, Q_1). Table 3.1 lists all the possibilities.

Multiplying the conditional values of $f_{n+1}(s, w)$ by their probabilities and summing over the rows of Table 3.1, we obtain the pgf recurrence:

$$f_{n+1}(s, w) = [p_2 f_n(s, w) + p_1 w + p_0]^2. \tag{3.13}$$

Let us note that if we limit ourselves to the proliferating cells, we obtain a GW process. Indeed, passing to the marginal pgf in Eq. (3.13), by setting $w = 1$, yields $f_{n+1}(s) = [p_2 f_n(s) + p_1 + p_0]^2$, which is a special case of Eq. (3.2) with $f(s) = (p_2 s + p_1 + p_0)^2$.

3.2.2 *Modeling Biological Data*

In the paper Kimmel and Axelrod (1991), data on the colonies of cells have been modeled with the aid of Eq. (3.1). The data included empirical distributions of colony sizes of two varieties of cultured mouse fibroblast (skin tissue) cells. The first variety, the NIH cells, are relatively "normal" cells. The second variety was created by transferring the mutated *ras* oncogene, implicated in some malignant tumors, into

Table 3.2 Colony size distribution: data and parameter estimates

Cell type	NIH	NIH (*ras*)
Data		
Duration of the experiment (h)	96	96
Number of colonies	52	45
Colony size (cells/colony)		
Minimum	10	8
Maximum	116	214
Median	33	70
Estimated parameters		
Number of divisions	8	8
Probability of death (p_0)	0.15	0.15
Probability of quiescence (p_1)	0.1	0

Fig. 3.3 Distributions of colony sizes for the NIH cells. *Squares* represent experimental data and *continuous lines* have been generated by the model, as described in the text. The model satisfactorily reproduces the distributions of the NIH cells colony sizes. (Source: Kimmel and Axelrod 1991)

NIH cells. The purpose of the experiments was to establish the differences in growth processes between the NIH and NIH (*ras*) cells.

The distributions of colony sizes, i.e., of the numbers of cells per colony, were obtained for a number of colonies grown for identical time in identical conditions. The wide variability of colony sizes demonstrates the utility of including a stochastic component in modeling. Table 3.2 provides a summary of data. Cumulative frequencies of colony sizes are depicted in Figs. 3.3 and 3.4

In the experiment, it is impossible to discern proliferative cells from quiescent cells. Therefore, a version of Eq. (3.13) is used, which gives the distributions of $Z_n + Q_n$. The pgf of this sum is equal to $g_n(s) = f_n(s, s)$, and therefore Eq. (3.13) yields

$$g_{n+1}(s) = [p_2 g_n(s) + p_1 s + p_0]^2. \tag{3.14}$$

Fig. 3.4 Distributions of colony sizes for the NIH (*ras*) cells. *Squares* represent experimental data and *continuous lines* have been generated by the model, as described in the text. The model satisfactorily reproduces the distributions of the NIH (*ras*) cells colony sizes. (Source: Kimmel and Axelrod 1991)

This pgf is equal to $g_n(s) = \sum_j \pi_n(j)s^j$, where $\pi_n(j) = \mathrm{P}\{Z_n + Q_n = j\}$, and Eq. (3.14) is equivalent to

$$\{\pi_{n+1}(j)\} = \{p_2\pi_n(j) + p_1\delta_{j1} + p_0\delta_{j0}\} * \{p_2\pi_n(j) + p_1\delta_{j1} + p_0\delta_{j0}\}, \quad (3.15)$$

where the asterisk denotes the convolution of distributions, i.e., $\{c(j)\} = \{a(j)\} * \{b(j)\}$ denotes $c(j) = \sum_{i=0}^{j} a(i)b(j-i)$, for $j = 0, 1, \ldots$. The Kronecker symbol, δ_{jk}, is equal to 1 if $j = k$, and equal to 0, if $j \neq k$. Recurrence (3.15), together with the condition $\pi_0(j) = \delta_{j1}$, $j \geq 0$, makes possible to calculate the distributions of colony sizes.

During the time of the experiment, about $n = 8$ divisions occurred. Distributions $\{\pi_8(j),\ j \geq 0\}$ of colony size can be computed using different values of probabilities p_0 and p_1 of cell death and quiescence. The following strategy has been used:

1. Since the NIH(*ras*) cells have increased content of the mutated *ras* oncogene product, they are not likely to be quiescent; therefore, $p_1 = 0$ is assumed and the probability p_0 of cell death is varied to fit the empirical distribution, in the sense of least sum of squares of deviations of the model from the data. This gives $p_0 = 0.15$ (Fig. 3.4).
2. For the NIH cells, which are "normal," the same value ($p_0 = 0.15$) of the probability of cell death is used, but the probability p_1 of quiescence is varied until the distribution fits the data. This gives $p_1 = 0.1$ (Fig. 3.3).

As evident from Fig. 3.3, the empirical cumulative frequency is initially steep, which suggests that colonies with less than 10 cells constitute a nonnegligible fraction of the sample. These colonies were not counted in the experiment, and therefore the theoretical distribution is calculated conditional on the event that the colony size is not less than 10.

The modified GW process accurately reproduces variability of the colony size.

3.3 Extinction and Criticality

In this section, we consider the classification into the subcritical, critical, and supercritical processes and the laws of process extinction. This material overlaps with Sect. 1.5.3, but it seems convenient to reintroduce it here.

The properties of the GW process are equivalent to the properties of the iterates $f_n(s)$ of the progeny pgf $f(s)$. In particular, the asymptotic behavior of $\{f_n(s)\}$ provides insight into the limit theorems for the $\{Z_n\}$ process.

Let s be a real number. From the definition of $f(s)$ as a power series with nonnegative coefficients $\{p_k\}$ and with $p_0 + p_1 < 1$, we see that

1. $f(s)$ is strictly convex and increasing in [0,1],
2. $f(0) = p_0$; $f(1) = 1$,
3. if $m \leq 1$, then $f(s) > s$ for $s \in [0,1)$,
4. if $m > 1$, then $f(s) = s$ has a unique root in $[0, 1)$.

Let q be the smallest root of $f(s) = s$ for $s \in [0, 1]$. Then the above properties imply that there is such a root and furthermore:

Lemma 3.1 *If $m \leq 1$ then $q = 1$; if $m > 1$ then $q < 1$.*

The properties stated before and in the Lemma 3.1 are easy to understand if a graph of the pgf is drawn. Moreover, we can prove that the iterates of $f(s)$ converge to q.

Lemma 3.2 *If $s \in [0, q)$ then $f_n(s) \uparrow q$ as $n \to \infty$. If $s \in (q, 1)$ then $f_n(s) \downarrow q$ as $n \to \infty$. If $s = q$ or 1 then $f_n(s) = s$ for all n.*

As a special case of the Lemma 3.2 we note that $f_n(0) \uparrow q$. But

$$\begin{aligned} \lim_{n\to\infty} f_n(0) &= \lim_n P\{Z_n = 0\} = \lim_n P\{Z_i = 0, \text{for some} 1 \leq i \leq n\} \\ &= P\{Z_i = 0, \text{for some } i \geq 1\} = P\{\lim_{n\to\infty} Z_n = 0\}, \end{aligned}$$

which is, by definition, the probability that the process ever becomes extinct. Applying Lemma 3.1, we get the extinction probability theorem.

Theorem 3.1 *The extinction probability of the $\{Z_n\}$ process is the smallest nonnegative root q of the equation $s = f(s)$. It is equal to 1 if $m \leq 1$, and it is less than 1 if $m > 1$.*

Theorem 3.1 states that the extinction probability depends on the parameter m, the mean progeny number of a particle.

Definition 3.1 *If m is less than 1, equal to 1, or greater than 1, then the process is called subcritical, critical, and supercritical, respectively.*

According to Theorem 3.1, the subcritical and critical processes become eventually extinct with probability 1. This is particularly surprising in the case of the critical process, for which, the expected value of $\{Z_n\}$ stays constant. Therefore, some branching process models behave differently from their deterministic counterparts. Early in the history of the branching processes, it was remarked (Harris 1963),

and then reiterated (Athreya and Ney 2004) that this instability of the GW process is contrary to the behavior of biological populations, which tend to reach a state of balance with their environment. We will see, based on examples taken from cell and molecular biology, that the phenomena of extinction and instability do not contradict the rules of biology.

The GW process is a Markov chain, the state of which is equal to the number of particles present. We may classify its states into *transient* and *recurrent.* Recurrent states are revisited with probability 1. For transient states, this probability is less than 1. Let us examine the probability of not returning to a given state k. Let us denote by $P(k, j) = P\{Z_{n+1} = j | Z_n = i\}$, the one-step transition probability of the process. The following is obtained,

$$P\{Z_{n+i} \neq k, \text{ for all } i \geq 1 | Z_n = k\} \geq \left\{ \begin{array}{ll} P(k,0) = p_0^k; & p_0 > 0 \\ 1 - P(k,k) = 1 - p_1^k; & p_0 = 0 \end{array} \right\} > 0. \tag{3.16}$$

The above is demonstrated as follows. If $p_0 > 0$, then one of the ways of not returning to state k is that the process becomes extinct in one step (this occurs with probability p_0^k). If this is impossible, i.e., if $p_0 = 0$, then we notice that one of the ways of not returning to k is to not return in a single step (if $p_0 = 0$, this occurs with probability $1 - p_1^k$). From Eq. (3.16) we deduce the following theorem:

Theorem 3.2 *All states except* $\{Z_n = 0\}$ *are transient, i.e.,*

$$P\{Z_{n+i} \neq k, \textit{ for all } i \geq 1 | Z_n = k\} > 0,$$

if $k \neq 0$. *In particular, this implies* $\lim_{n\to\infty} P\{Z_n = k\} = 0$ *and* $P\{\lim_{n\to\infty} Z_n = k\} = 0$, *for* $k \geq 1$. *The above, together with Theorem 3.1, implies that*

$$P\{\lim_{n\to\infty} Z_n = 0\} = 1 - P\{\lim_{n\to\infty} Z_n = \infty\} = q.$$

This latter property is known as the instability of the branching process.

3.4 Application: Complexity Threshold in the Evolution of Early Life

This example is taken from a paper by Demetrius et al. (1985). It concerns the ability of long biomolecules (*polymeric chains*) composed of smaller units (*nucleotides*) to replicate without error. The same problem can be considered in many different frameworks.

Let us consider a polymeric chain of ν nucleotides. If we assume that there is a fixed probability p that a single nucleotide is correctly copied, then the probability that a copy of the whole chain is correct is p^ν. Let us suppose that the chain replicates

in a single time unit. During one-generation step, the molecule either survives (with probability w) and produces a copy, which is accurate with probability p^ν, or it is destroyed with probability $1 - w$. A given molecule yields 0, 1, or 2 molecules of the same type after one unit of time: the probabilities are $1 - w$, $w(1 - p^\nu)$, and wp^ν, respectively. The population of error-free molecules evolves according to a GW branching process with pgf $f(s) = (1 - w) + w(1 - p^\nu)s + wp^\nu s^2$. This biomolecule is indefinitely preserved with a positive probability only if the process is supercritical, i.e., if $m = w(1 + p^\nu) > 1$, which yields

$$p^\nu > \frac{1 - w}{w}. \tag{3.17}$$

The probability of nonextinction is equal to $1 - (1 - w)/(wp^\nu)$.

Relation (3.17) implies that the error probability $1 - p$ implies a threshold for the length ν of the molecule, which does not become extinct with probability 1. If ν is larger than this threshold, then the molecular species becomes extinct.

3.5 Asymptotic Properties

The limit theorems for the GW process are important for many applications. Also, they suggest what to expect in more complicated processes. The limit laws are different in the supercritical, subcritical, and critical cases. In the supercritical case, the principal result is that the growth is asymptotically exponential, and that with probability 1 the random variable Z_n/m^n tends to a limit W. As a consequence, Z_n is approximated by Wm^n for large n. This is an extension of the exponential or Malthusian law of growth in the realm of stochasticity.

However, here the analogy ends. In the subcritical and critical cases, the probability of extinction is equal to 1 and the limit of Z_n/m^n is 0. Therefore, the "Malthusian law" is no longer a sensible approximation. In the subcritical case, it is replaced by the limit laws conditional on nonextinction, i.e., for the process $\{Z_n | Z_n > 0\}$. In the critical case, the limit distribution of $\{\frac{Z_n}{n} | Z_n > 0\}$ is exponential.

3.5.1 Supercritical Process

The main mathematical fact used in this case is that the process $\{Z_n/m^n\}$ is a martingale.

Definition 3.2 *A sequence of random variables $\{X_n, n \geq 0\}$ is called a martingale if $E(|X_0|) < \infty$, and*

$$E(X_{n+1} | X_n, X_{n-1}, \ldots, X_1, X_0) = X_n.$$

Theorem 3.3 *If $\{X_n, n \geq 0\}$ is a nonnegative martingale, such that $E(X_n) < \infty$ for all n, then there exists a proper random variable X with finite expectation such that*

(i)

$$\lim_{n\to\infty} X_n = X, \text{ wp } 1.$$

(ii) *If the martingale is L^2 bounded, i.e., if $\sup_n E(X_n^2) < \infty$, then the convergence occurs also in the L^2 sense. Then, $Var(X) = \lim_{n\to\infty} Var(X_n)$.*

Theorem 3.3 is a modified form of the theorem in Sect. 1.3 of the book by Neveu (1975).

If we set

$$W_n \equiv Z_n/m^n,$$

then $\mathrm{E}(W_n) = 1$ and

$$\mathrm{E}(W_{n+1}|W_n) = m^{-(n+1)}\mathrm{E}(Z_{n+1}|Z_n) = m^{-(n+1)}mZ_n = W_n. \tag{3.18}$$

Since the GW process is a Markov chain, then

$$E(Z_{n+1}|Z_n, Z_{n-1}, \ldots, Z_1, Z_0) = E(Z_{n+1}|Z_n).$$

An analogous property holds for the normalized process $\{W_n\}$, i.e.,

$$E(W_{n+1}|W_n, W_{n-1}, \ldots, W_1, W_0) = E(W_{n+1}|W_n).$$

Consequently, Eq. (3.18) demonstrates $\{W_n\}$ is a martingale. Therefore, by part (i) of Theorem 3.3 we obtain the following:

Theorem 3.4 *If $0 < m \equiv f'(1-) < \infty$, then there exists a random variable W such that*

$$\lim_{n\to\infty} W_n = W, \text{ wp } 1.$$

In the critical and subcritical case, $W \equiv 0$ since $q = 1$. Therefore, W might be nondegenerate only if $m > 1$. This is indeed true if an additional condition of finite variance of the number of progeny is imposed.

Theorem 3.5 *If $m > 1$, $\sigma^2 < \infty$, and $Z_0 \equiv 1$, then (i) $\lim_{n\to\infty} E(W_n - W)^2 = 0$; (ii) $E(W) = 1, Var(W) = \sigma^2/(m^2 - m)$; (iii) $P\{W = 0\} = q = P\{Z_n = 0$ for some $n\}$.*

The Laplace transform $\phi(v) = \mathrm{E}(\mathrm{e}^{-vW})$ of the distribution of W can be shown to satisfy a functional equation,

$$\phi(v) = f\left[\phi\left(\frac{v}{m}\right)\right], \tag{3.19}$$

the so called Abel's equation. Indeed, the relationship (3.2) can be rewritten in the terms of Laplace transforms $\varphi_n(u) = \mathrm{E}[\exp(-uZ_n)] = f_n[\exp(-u)]$, where $u \geq 0$ is a symbolic argument,

$$\varphi_{n+1}(u) = f[\varphi_n(u)]. \tag{3.20}$$

Since the Laplace transform of the distribution of W_n is equal to

$$\phi_n(u) = E[\exp(-uW_n)] = E\left[\exp\left(-\frac{u}{m^n}Z_n\right)\right] = \varphi_n(u/m^n)$$

and conversely $\varphi_n(u) = \phi_n(um^n)$, substitution into Eq. (3.20) yields $\phi_{n+1}(um^{n+1}) = f[\phi_n(um^n)]$. After change of variables $v = um^{n+1}$, we obtain

$$\phi_{n+1}(v) = f\left[\phi_n\left(\frac{v}{m}\right)\right].$$

Since $W_n \to W$ in distribution, then $\phi_n(v) \to \phi(v)$ and the limit (which is the Laplace transform of the distribution of rv W) satisfies the Abel's Eq. (3.19).

Example: In the linear-fractional case of Sect. 3.1.4, the distribution of W can be directly calculated. Its "density" can be expressed as,

$$f_W(w) = q\delta(w) + (1-q)^2 \mathrm{e}^{-(1-q)w}, \ w \geq 0,$$

i.e., it has an atom at 0 and the remaining part is negative exponential (c.f., Problems at the end of this chapter).

3.5.2 Subcritical Process

In the subcritical case, the process becomes extinct with probability 1. What can be said about the asymptotic behavior?

Example: Linear-fractional case. The probability of nonextinction is now equal to

$$1 - f_n(0) = m^n\left(\frac{1-q}{m^n - q}\right),$$

(c.f., Problems at the end of this chapter) which yields

$$\mathrm{E}(Z_n | Z_n > 0) = \frac{\mathrm{E}(Z_n)}{1 - f_n(0)} = \frac{m^n - q}{1-q} \to \frac{q}{q-1}; \quad n \to \infty.$$

This suggests that conditioning on nonextinction is a sufficient device to obtain a limit law. The proof of the following result can be found in the book of Athreya and Nei (2004):

Theorem 3.6 *Yaglom's. If $m < 1$ then $P\{Z_n = j | Z_n > 0\}$ converges, as $n \to \infty$, to a probability function whose generating function $\mathcal{B}(s)$ satisfies equation*

$$\mathcal{B}[f(s)] = m\mathcal{B}(s) + (1-m). \tag{3.21}$$

Also,

$$1 - f_n(0) \sim \frac{m^n}{\mathcal{B}'(1-)}, \quad n \to \infty. \tag{3.22}$$

The above theorem will be useful in the next application considered in this chapter. Convergence to a limit distribution conditional on nonabsorption is known as quasistationarity (see Sects. 7.4 and 7.6).

3.5.3 Critical Process

By analogy, to the deterministic case it might appear that in the critical case, in which $W_n = Z_n$, the sequence Z_n might reach a nontrivial limit. However, it is impossible, since the extinction probability is equal to 1 in this case. To approximate the asymptotic behavior of the critical GW process, it is necessary to use conditioning on nonextinction and normalization. Let us start with a basic lemma (Athreya and Ney 2004).

Lemma 3.3 *If* $m = E(Z_1) = 1$ *and* $\sigma^2 = Var(Z_1) < \infty$, *then*

$$\lim_{n\to\infty} \frac{1}{n}\left[\frac{1}{1 - f_n(t)} - \frac{1}{1-t}\right] = \frac{\sigma^2}{2}$$

uniformly for $0 \le t < 1$.

Linear-fractional case: If $m = 1$, then $\mathrm{Var}(Z_1) = f''(1-) = 2\ p/(1-p) < \infty$, based on Eq. (3.10). In this case, the assertion of lemma is obtained by directly computing the limit (c.f., Problems at the end of this chapter).

Based on the lemma, the rate at which the critical process becomes extinct can be estimated. The limit behavior of the probability of *nonextinction*, $\mathrm{P}\{Z_n > 0\}$, is found by setting $t = 0$ in the Lemma 3.3

$$\mathrm{P}\{Z_n > 0\} = 1 - f_n(0) \sim \frac{2}{n\sigma^2}; \quad n \to \infty.$$

By the same token, the expectation of the process, given nonextinction, satisfies

$$\mathrm{E}(Z_n | Z_n > 0) = \frac{1}{\mathrm{P}\{Z_n > 0\}} \sim \frac{n\sigma^2}{2}; \quad n \to \infty.$$

This latter suggests that a limit law could exist for the normalized and conditional process $\left\{\frac{Z_n}{n} | Z_n > 0\right\}$. Indeed, we have,

Theorem 3.7 *If* $m = 1$ *and* $\sigma^2 < \infty$, *then*

$$\lim_{n\to\infty} \mathrm{P}\left\{\frac{Z_n}{n} > z \middle| Z_n > 0\right\} = \exp\left\{-\frac{2z}{\sigma^2}\right\}, \ z \ge 0.$$

For the proof, see Athreya and Ney (2004).

3.6 Application: Cancer Mutations

In past several years, a number of interesting models of mutations leading to cancer have been published. They all explore models of proliferation, frequently using branching processes, combining them with models of driver and passenger mutations. Driver mutations are those that, although they might have arisen spontaneously, provide selective advantage for the emerging cancer proliferation, particularly against the background of already existing inherited or acquired mutations. Passenger mutations are generally neutral and their accumulation may provide a molecular "clock" indicating how long it has been since the cancer cells deviated from normal cells.

3.6.1 Modeling Driver and Passenger Mutations

A mathematical model of the relationship between accumulation of driver and passenger mutation in tumors was published by Nowak's group (Bozic et al. (2010)). The model in that paper is based on the GW branching process. The hypotheses are as follows: At each time step, a cell can either divide or differentiate, senesce, or die. In the context of tumor expansion, there is no difference between differentiation, death, and senescence, because none of these processes will result in a greater number of tumor cells than present prior to that time step. It is assumed that driver mutations reduce the probability that the cell will become "stagnate," i.e., that it will differentiate, die, or senesce, although the stagnant cells are not removed from the tumor. A cell with k driver mutations has a stagnation probability $d_k = (1-s)^k/2$. The division probability is $b_k = 1 - d_k$. The parameter s is the selective advantage provided by a driver mutation. When a cell divides, one of the daughter cells can receive an additional driver mutation with probability u. The theory can accommodate any realistic mutation rate and the major numerical results are only weakly affected by varying the mutation rate.

We can calculate the average time between the appearance of successful cell lineages. Not all new mutants are successful, because stochastic fluctuations may lead to the extinction of a lineage. The lineage of a cell with k driver mutations survives only with a probability of approximately $1 - d_k/b_k \cong 2sk$. Assuming that $u \ll ks \ll 1$, the average time between the first successful cell with k and the first successful cell with $k+1$ driver mutations is given by

$$\tau_k = \frac{T}{ks} \log \frac{2ks}{u}. \tag{3.23}$$

This result is obtainable from the theory of the GW process (by elementary means) and the derivation is found in the supplement to Bozic et al. (2012). The cumulative time to accumulate k mutations grows logarithmically with k. On the other hand, the average number of passenger mutations, $n(t)$, present in a tumor cell after t days is proportional to t, that is $n(t) = \nu t/T$, where ν is the rate of acquisition of neutral

mutations. Combining the results for driver and passenger mutations, results in a formula for the number of passenger mutations that are expected in a tumor that has accumulated k driver mutations

$$n = \frac{\nu}{2\,s} \log \frac{4ks^2}{u^2} \log k. \tag{3.24}$$

Here, n is the number of passengers that were present in the last cell that clonally expanded. Bozic et al. (2010) demonstrate that this dependence fits empirical data on several human cancers.

3.6.2 *Distribution of Mutational Events in Various Phases of Tumor Growth*

This question has been addressed by Tomasetti et al. (2013). The framework is not very different from that of Bozic et al. (2010). However, several models are considered, corresponding to different phases of proliferation in the tissue: precancerous (split into developmental and tissue renewal subphases) and cancerous phase. In the developmental phase, a branching process was used. In the tissue renewal phase, a Moran model was used. In the cancerous phase, a branching process was used again.

Together, the models make the novel prediction, validated by empirical findings, that the number of somatic mutations in tumors of self-renewing tissues is positively correlated with the age of the patient at diagnosis. Importantly, the analysis indicates that half or more of the somatic (new, since birth) mutations in tumors of self-renewing tissues occur prior to the onset of neoplasia. The model also provides a novel way to estimate the in vivo tissue-specific somatic mutation rates in normal tissues directly from the sequencing data of tumors (Fig. 3.5).

Gupta et al. (2011) consider the question of distinct phenotypic states of cells in tumors that differ in functional attributes. The mechanism by which the proportions of cells in different states are stabilized, which is not well understood, is studied in this paper using stochastic transitions occurring according to a Markov model. The biological example used is human breast cancer cell lines. It is shown experimentally and verified by a Markov process model that if the subpopulations are mixed in nonequilibrium proportions, then the mixture will gradually return to equilibrium. This is interesting although the mathematical model used is very simplistic. One might hypothesize that in growing cell populations, which can be modeled by a supercritical multitype branching process, a similar phenomenon takes place, which is predicted if the transition matrix is irreducible.

Komarova and Cheung (2005) employ a reducible multitype branching process (called "finite branching process" in the paper) to describe proliferation and maturation of cells in the colon crypt. The idea is to find, theoretically, the number of

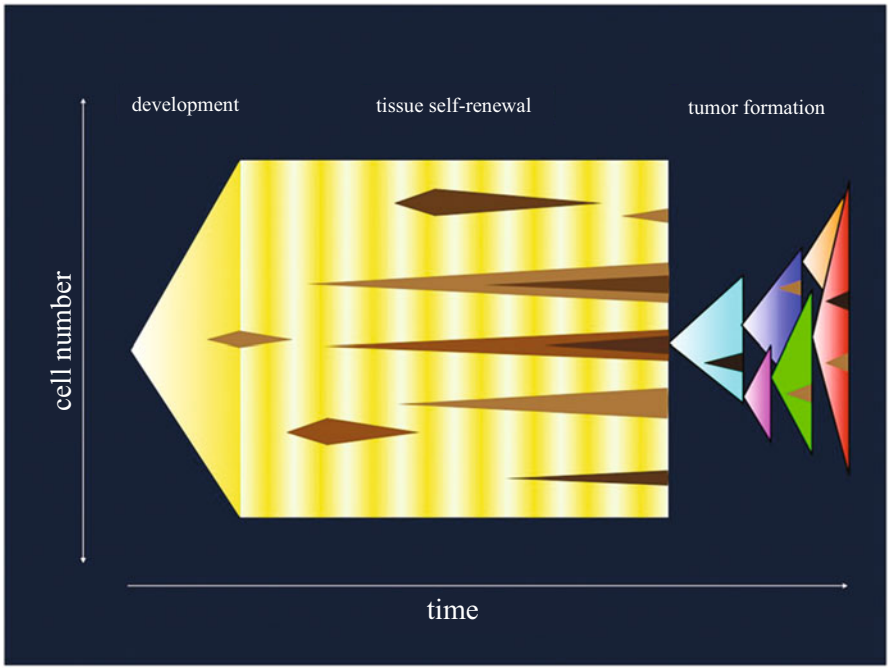


Fig. 3.5 The fish, a schematic representation of the different phases in which somatic mutations occur in a tissue giving rise to a cancer. Starting from a single precursor cell, a tissue is created via clonal expansion (head of the fish). The tissue is then subjected to periodic self-renewals (body of the fish). During development and tissue renewal, passenger mutations occur randomly, undergo clonal expansions (various brown clones), and either go extinct or expand as successive passenger mutations accumulate. A driver gene mutation can initiate a tumor cell clone, which then can expand through subsequent driver mutations, eventually yielding a clinically detectable tumor mass (fish's tail, where each clonal expansion driver by a new driver mutation is indicated by a different color). Passenger mutations occur during this phase as well. (Source: Tomasetti et al. 2013)

maturation stages and other parameters that minimize the probability that malignant clone will be established in the crypt. The conclusion is that the process in the crypts has evolved in an optimum manner.

3.7 Application: Gene Amplification

Material of this section is based on the paper by Kimmel and Axelrod (1990). It is an example of application of the Yaglom's Theorem 3.6 to the analysis of the asymptotic behavior of a subcritical GW process.

3.7.1 Gene Amplification and Drug Resistance

Amplification of a gene is an increase of the number of copies of that gene in a cell. Amplification of genes coding for the enzyme dihydrofolate reductase (DHFR) has been associated with cellular resistance to the anticancer drug methotrexate (MTX).

A resistant population with an increased number of DHFR gene copies per cell can be obtained after a sensitive population is grown in increasing concentrations of the drug. Increased resistance is correlated with increased numbers of gene copies on small extrachromosomal DNA elements. These elements are visible in the microscope and resemble pairs of small chromosomes; they are called double minute chromosomes or *double minutes*. The number of DHFR genes on double minutes in a cell may increase or decrease at each cell division. This is because double minutes are acentric, i.e., they do not have centromeres like real chromosomes. Centromeres are required for the mitotic apparatus to faithfully segregate chromosomes into progeny cells.

In populations of cells with the double minutes, both the increased drug resistance and the increase in number of gene copies are reversible. The classical experiment confirming this includes transferring the resistant cell population into a drug-free medium. When these populations are grown in the absence of the drug, they gradually lose resistance to the drug, by losing extra gene copies.

The population distribution of numbers of copies per cell can be estimated by the experimental technique called *flow cytometry*. In the experiments described, two features of these distributions are notable. First, as expected, the proportion of cells with amplified genes decreases with time. Second, less obvious, the shape of the distribution of gene copy number within the subpopulation of cells with amplified genes appears stable as resistance is being lost. This stable distribution is depicted in Fig. 3.6 taken from Brown et al. (1981). The distribution of cells with amplified genes retains its shape, only the area under the distribution gradually decreases while the peak corresponding to sensitive cells increases.

3.7.2 Galton–Watson Process Model of Gene Amplification and Deamplification

We consider a cell, one of its progeny (randomly selected), one of the progeny of that progeny (randomly selected), and so forth. The cell of the nth generation contains Z_n double minutes carrying the DHFR genes. During cell's life, each double minute is either replicated with probability a, or not replicated, with probability $1 - a$, independently of the other double minutes. Then, at the time of cell division, the double minutes are segregated to progeny cells. If the double minute has not been replicated, then it is assigned to one of the progeny cells with probability $\frac{1}{2}$. If it has been replicated, then either both copies are assigned to progeny 1 (wp $\alpha/2$), or to progeny 2 (wp $\alpha/2$), or they are divided evenly between both progeny (wp $1-\alpha$). Let

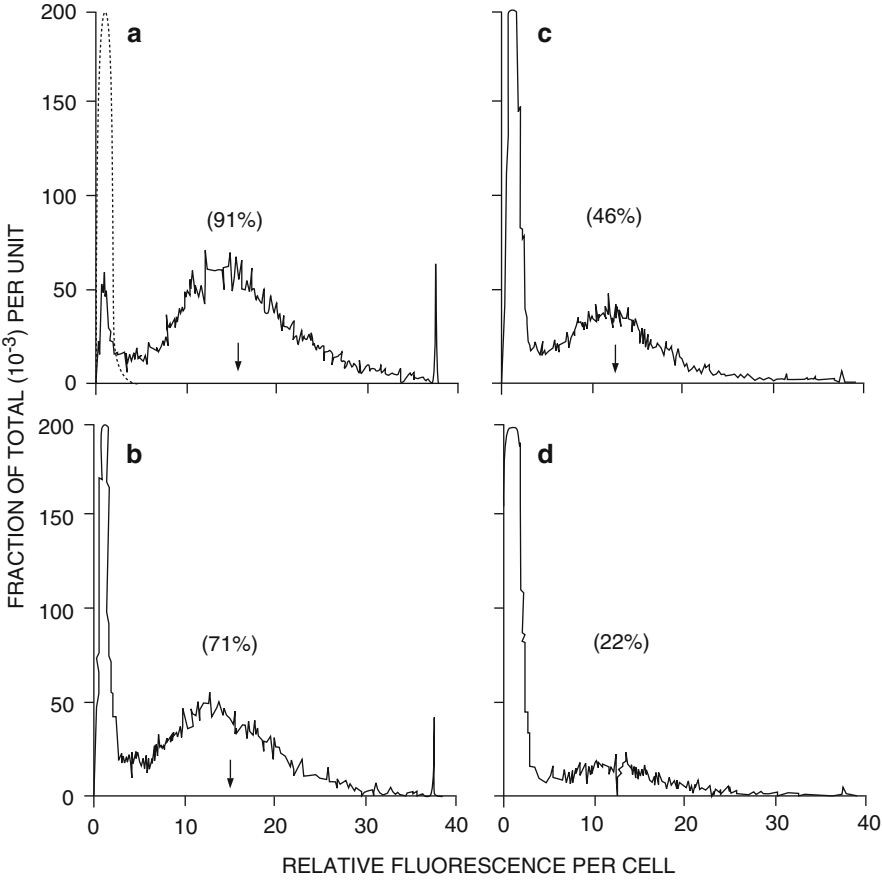


Fig. 3.6 Loss of the amplified copies of the DHFR gene during cell growth in MTX-free media. The 3T6 cells resistant to the MTX were grown for different times in MTX-free medium. The fluorescence level is proportional to the number of gene copies per cell. The values in parentheses are the percentages of cells with gene copy numbers greater than those for sensitive cells. **a** Dotted line, 3T6 sensitive cells; solid line, resistant cells. **b** Cells grown for 17 generations without MTX. **c** Cells grown for 34 generations without MTX. **d** Cells grown for 47 generations without MTX. (Source: Brown et al. 1981)

us note that the two double minutes segregate independently to progeny cells only when $\alpha = 1/2$. Otherwise, they either preferentially go to the same cell ($\alpha > 1/2$), or to different cells ($\alpha < 1/2$). The randomly selected progeny in our line of descent contains (Fig. 3.7),

- No replicas of the original double minute (wp $(1 - a)/2 + a\alpha/2$), or
- One replica of the original double minute (wp $(1 - a)/2 + a(1 - \alpha)$), or
- Both replicas of the original double minute (wp $a\alpha/2$).

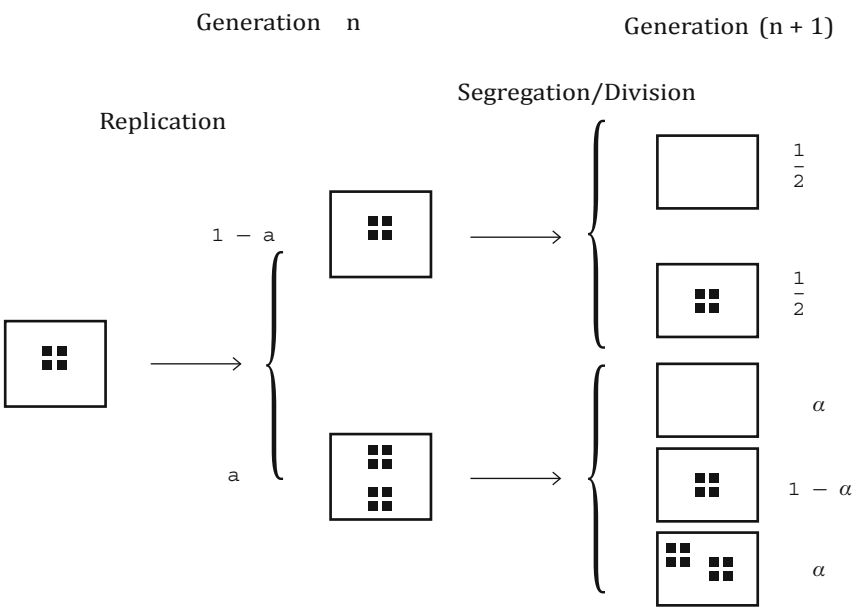


Fig. 3.7 Schematic representation of the mathematical model of amplification and deamplification of genes located on double minute chromosomes. The sequence of events is presented for one of the possibly many double minutes present in the cell. During cell's life, the double minute is either replicated, or not replicated. At the time of cell division, the double minute is assigned to one of the daughter cells (*segregation*). If it has not been replicated, it is assigned to one of the daughter cells. If it has been replicated, then either both copies are assigned to daughter 1, or to daughter 2, or they are divided evenly between both daughters. Probabilities of the events involved are presented in the graph. (Source: Kimmel and Axelrod 1990)

Therefore, the number of double minutes in the nth generation of the cell lineage is a GW process with the progeny pgf

$$f(s) = d + (1 - b - d)s + bs^2, \tag{3.25}$$

where $b = a\alpha/2$ and $d = (1-a)/2+a\alpha/2$ are the probabilities of *gene amplification* and *deamplification*, respectively. Since in the absence of selection double minutes gradually disappear from the cell population, it is assumed that deamplification (loss of gene copies) exceeds amplification, so that the process is subcritical. In mathematical terms, $b < d$ and $m = f'(1-) = 1 + b - d < 1$.

3.7.3 Mathematical Model of the Loss of Resistance

We will call a cell *resistant* if it carries at least one double minute chromosome with the DHFR gene. Otherwise it is called *sensitive*. In the experiments described

above, a population of cells resistant to MTX, previously cultured for N generations in medium containing MTX, consists only of cells with at least one DHFR gene copy, i.e., $Z_N > 0$. Therefore, the number of gene copies per cell is distributed as $\{Z_N | Z_N > 0\}$. If N is large then, since the process is subcritical, by the Yaglom Theorem 3.6, this distribution has pgf $\mathcal{B}(s)$ satisfying the functional equation given in the theorem. Also, based on the estimates of $1 - f_n(0)$ provided in the same theorem, the resistant clone grows, in each generation, by the factor $2m$ on the average.

After the N initial generations, the resistant clone has been transferred to the MTX-free medium. The *overall* number of cells now grows by factor two in each generation, while the average number of *resistant* cells continues to grow by the factor $2m$. Let us denote by $R(n)$ and $S(n)$, the number of resistant and sensitive cells in the population, n generations after transferring the cells to the MTX-free medium; $r(n) = R(n)/[R(n) + S(n)]$ is the fraction of resistant cells. We obtain,

$$R(n) = (2m)^n R(0),\ S(n) + R(n) = 2^n [S(0) + R(0)],$$

hence

$$r(n)/r(0) = m^n. \tag{3.26}$$

This means that the proportion of resistant cells decreases geometrically, while the distribution of gene copy number among the resistant cells remains close to the limit distribution of the Yaglom theorem. This behavior is consistent with the experimental data of Fig. 3.6.

3.7.4 Probabilities of Gene Amplification and Deamplification from MTX Data

Probabilities b and d can be estimated from the loss of resistance experiments similarly as in Kimmel and Axelrod (1990), using data on the S-180(R_1A) cells in Kaufman et al. (1981). The resulting estimates are $b = 0.47, d = 0.50$, yielding $a = 1 - 2(d - b) = 0.94$ and $\alpha = 2b/a = 1$. The interpretation is that while the frequency of replication of the double minute chromosomes is quite high, both copies are assigned almost always to the same progeny cell.

In Kimmel and Axelrod (1990), other models of the same process have been considered. All of them exhibit dynamics similar to that predicted by the Yaglom theorem.

3.8 Application: Iterated Galton–Watson Process and Expansion of DNA Repeats

We consider mathematical properties of a time-discrete stochastic process describing explosive proliferation of DNA repeats in a class of human genetic diseases. The process contains copies of the GW process as its building blocks.

3.8.1 Dynamics of DNA Repeats in Human Pedigrees

Recently, several heritable disorders have been associated with dynamic increases of the number of repeats of DNA triplets in certain regions of human genome. In two to three subsequent generations, the transitions from normal individuals to non-affected or mildly affected carriers, and then to full-blown disease, occur. The two syndromes for which the most comprehensive data exist are (Richards and Sutherland 1994):

- The Fragile X syndrome, caused by a mutation of the FMR-1 gene characterized by expansion of the $(CCG)_n$ repeats (normal 6–60, carrier 60–200, affected >200 repeats).
- Myotonic dystrophy, caused by a mutation of the DM-1 autosomal gene characterized by expansion of the $(AGC)_n$ repeats (normal 5–27, affected >50 repeats).

These two inherited human syndromes previously were distinguished by two features inconsistent with Mendelian inheritance: progressively earlier onset of symptoms in subsequent generations and higher severity of symptoms in subsequent generations.

These features have recently been correlated with changes in DNA. In each case, a trinucleotide represented a few times in an unaffected parents is found in multiple tandem copies in the progeny. The number of tandem copies is dramatically increased (10–100-fold) in affected individuals. It has been correlated with the time of onset and the severity of symptoms.

Important questions that have not been fully answered are:

1. What is the mechanism of fluctuation of the number of repeat sequences in normal people (not in affected families)?
2. What is the mechanism of the modest increase in repeat sequences in unaffected carriers?
3. What is the mechanism of the rapid expansion of the number of repeat sequences in affected progeny within one or two generations?

Caskey et al. (1992) formulated a biological hypothesis regarding the origin of high variation in repeat count:

> The instability in the premutation alleles which leads to the extraordinary expansions observed in DM and Fragile X patients results from the presumed difficulty of replicating long GC-rich sequences. In this scenario, unequal rates of DNA synthesis lead to multiple incomplete strands of complementary, triplet, reinitiated sequences.

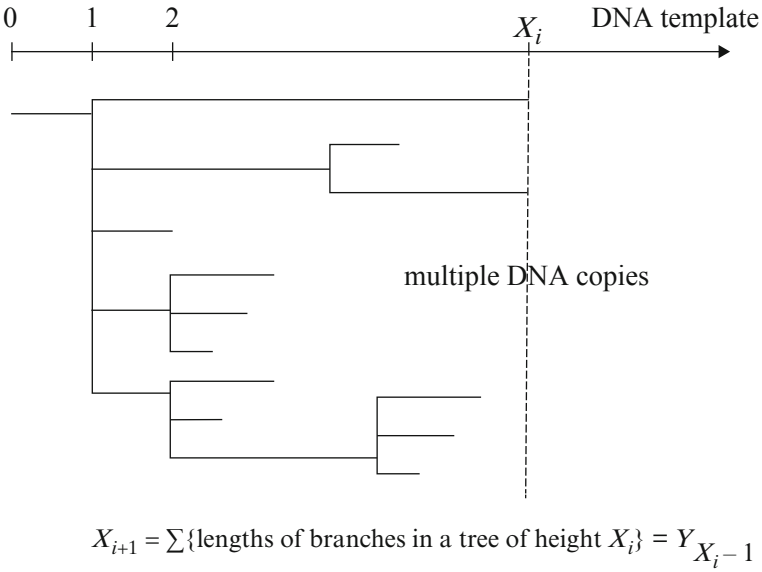


Fig. 3.8 The nonlinear mechanism of repeat expansion: Illustration of DNA branches that can be resolved into repeats. Suppose that in generation i there are X_i triplets and suppose that a random number (usually, 0 or 1) of new branches of DNA emerge on top of a previous one at the endpoint of each repeat (a single "initiation before termination" on each branch). Assuming that the process is confined to a region defined by the length of the original X_i repeats, and that all triplets from all branches are resolved and incorporated into a linear structure of chromosomal DNA, we obtain the number of repeats X_{i+1} in the $(i + 1)$st generation equals to the number of all triplets encased in the thin-line rectangle. (Source: Gawel and Kimmel 1996)

3.8.2 Definition of the Process

Gawel and Kimmel (1996) make this hypothesis specific by assuming the following specific scenario of expansion of repeats:

- In the initial, zeroth, replication round, the number of repeats is n.
- In each new DNA replication round, a random number of new branching events (i.e., "initiation without termination of replication" events) occur at the endpoint of each repeat (this random number is characterized by the pgf $f(s)$) and
- All resulting branches become resolved and reintegrated into the linear DNA structure, which becomes the template for the succeeding replication round.

Gawel and Kimmel (1996) notice that there exist a precedent for such mechanism in the replication of the T4 bacteriophage. This virus induces production in the host cell of branched networks of concatenated DNA, which subsequently is resolved into unbranched phage genomes (see references in Gawel and Kimmel 1996), Fig. 3.8.

Gawel and Kimmel (1996) developed the so-called iterated simple branching process $\{X_i\}$ to provide a mathematical formulation for the expansion process. Here

X_i is the length of a linear chain of DNA repeats after the ith stage of replication ($i = 0, 1, 2, \dots$) and $X_0 = n > 1$. A chain with $X_i = \nu$ repeats replicates as a branched network, which is assumed to be a GW tree descended from a single ancestor through $\nu - 1$ generations. Thus, the replicating chain serves as a template for the height of the daughter tree. This partial tree later resolves into a linear chain. To compute the length of this chain, let us suppose that

$$\{Z_k, k \geq 0\} \tag{3.27}$$

is the sequence of numbers of individuals in a GW process with progeny pgf $f(s)$. Suppose further that the sequence

$$\{Y_k, k \geq 0\}, \tag{3.28}$$

where

$$\begin{aligned} Y_0 &= Z_0 = 1, \\ Y_1 &= Z_0 + Z_1, \\ Y_2 &= Z_0 + Z_1 + Z_2, \\ &\dots \quad \dots \dots \dots \\ Y_k &= Z_0 + Z_1 + \cdots + Z_k, \\ &\dots \quad \dots \dots \dots \end{aligned} \tag{3.29}$$

is the total progeny process, i.e., Y_k is the cumulative number of progeny produced in the generations 0 through k of the GW process (c.f., Problems at the end of this chapter).

Let further $\{Z_k^{(i)}, k \geq 0\}$, $i \geq 0$ be a sequence of iid copies of $\{Z_k\}$ with $\{Y_k^{(i)}, k \geq 0\}$, $i \geq 0$ being the corresponding total progeny processes. These are the tree structures grown at each (ith) replication round. The generic process $\{Z_k\}$ is called the underlying GW process.

The process

$$\{X_i, i \geq 0\}, \tag{3.30}$$

can be now defined in a recursive manner,

$$X_0 = n, \tag{3.31}$$

$$X_{i+1} = Y_{X_i - 1}^{(i)}, \; i \geq 0. \tag{3.32}$$

Hence the sequence $\{X_i\}$ is a Markov process and, since $Y_0^{(i)} = 1$, the state 1 is absorbing.

3.8.3 Example

The following version of the process seems to be realistic from the biological viewpoint: Suppose that at the end of each repeat a new "initiation before termination" event occurs with small probability p, so that

$$f(s) = (1 - p)s + ps^2. \tag{3.33}$$

Then, the number of branches stemming from each ramification point is at least 1 and at most 2, the latter event being less likely. This leads to a "sparse" tree and implies that for a number of generations the growth of the process will be slow.

Fluctuations of the number of triplets in the unaffected individuals can be explained by coexistence of processes of triplet increase and triplet loss. Accordingly, we also assume that the process of resolution and reincorporation of repeats into the linear chromosomal structure has a limited efficiency $u < 1$.

This can be mathematically formalized using the idealized *binomial thinning*, i.e., assuming that each repeat is resolved and reinserted with probability u. The new process $\{\tilde{X}_i, i \geq 0\}$, including the imperfect efficiency is defined as

$$\tilde{X}_0 = n, \tag{3.34}$$

$$\tilde{X}_{i+1} = B(u, Y^{(i)}_{\tilde{X}_i - 1}) + 1,\ i \geq 0, \tag{3.35}$$

where conditional on N, $B(u, N)$ is a binomial random variable with parameters u and N.

With an appropriate choice of parameters (see further on), this process produces runs of fluctuations, followed by an explosive growth.

3.8.4 Properties

For the process without thinning, Pakes (2000) provides the following analysis, which is simpler than the original arguments in Gawel and Kimmel (1996): We exclude the trivial case $p_1 = 1$, where $X_i = X_0$. Then $\mathrm{P}[\{X_i \to 1\} \cup \{X_i \to \infty\}] = 1$. Let X_∞ denote the almost sure limit of X_i and let $g(s, v)$ denote the pgf of Y_v. Then $g(s, 0) = s$ and $g(s, v + 1) = sf[g(s, v)]$ (see problems to the present chapter). It follows from (3.35) that

$$\mathrm{E}(s^{X_{i+1}}) = \mathrm{E}[g(s, X_i - 1)],$$

and hence in all cases

$$\mathrm{E}(s^{X_\infty}) = \mathrm{E}[g(s, X_\infty - 1)]. \tag{3.36}$$

If

$$0 < p_0 < 1, \tag{3.37}$$

we may choose $s \in (0, q)$ and then $f(s) > s$. This gives $g(s, 1) > sf(s) > s$ and hence, by induction, that $g(s, \nu - 1) > s^{\nu}$. Since (3.36) can be written as

$$s + \mathrm{E}(s^{X_\infty}, X_\infty > 1) = s + \mathrm{E}[g(s, X_\infty - 1), X_\infty > 1],$$

it is clear that this can hold if and only if $\mathrm{P}[X_\infty > 1] = 0$. We conclude that the process is absorbed at unity when (3.37) holds. Next, if $p_0 = 0$, then $Y_\nu^{(i)} > \nu + 1$ and hence (3.32) implies $X_{i+1} \geq X_i$. So, $X_i \uparrow \infty$ if $X_0 \geq 2$. The reasoning above (and some other details) are summarized by the following statement:

Theorem 3.8 *Let us consider the Iterated Galton-Watson (IGW) process with no thinning (i.e., with $u = 1$). Then*

1. *$m < 1$ yields $E(X_i) \to 1$ and $X_i \overset{a.s.}{\to} 1$.*
2. *$m = 1$ yields $E(X_i) = E(X_0)$ and $X_i \overset{a.s.}{\to} X_\infty$ where X_∞ is a finite rv, and $X_\infty = 1$ if $p_0 < 1$.*
3. *$m > 1$ yields $E(X_i) \to \infty$ and*
   a) if $p_0 > 0$, then $X_i \overset{a.s.}{\to} 1$,
   b) if $p_0 = 0$, i.e., $f(s) = p_1 s + p_2 s^2 + \cdots$, then $X_i \overset{p}{\to} \infty$.

The next result concerns the growth of the IGW process with binomial thinning.

Theorem 3.9 *Suppose $\{\tilde{X}_n\}$ is the IGW process with binomial thinning.*

1. *Suppose $m > 1$. For each integer $M > 0$, there exists an integer $N_0 > 0$ such that*

$$E(\tilde{X}_{i+1} | \tilde{X}_i = N_0) > MN_0.$$

2. *Suppose $u > 1/2$ and $p_0 = 0$. There exist $N_0 \geq 0$ and $\alpha > 1$ such that*

$$E(\tilde{X}_{n+1} | \tilde{X}_n \geq N_0) \geq \alpha E(\tilde{X}_n - 1 | \tilde{X}_n \geq N_0).$$

The properties stated in Theorem 3.8 are similar to those of the process, with the absorbing state being $\{X = 1\}$ in our case, as opposed to $\{X = 0\}$ for the GW process. The most notable difference is that the supercritical GW process never becomes absorbed with probability 1, while the iterated supercritical process may.

Theorem 3.9 shows that no matter how small the efficiency u in the process with thinning, the process will increase (in the expected value sense) by an arbitrary factor, only if it exceeded a certain threshold. To illustrate the properties of the process with thinning, 20 independent simulations with parameters $p = 0.05$ and $u = 0.8$ are presented in Fig. 3.9. All of them start from $n = 20$ repeats. Once the fluctuation exceeds 100–200 repeats, it usually jumps to $\geq$1000 repeats.

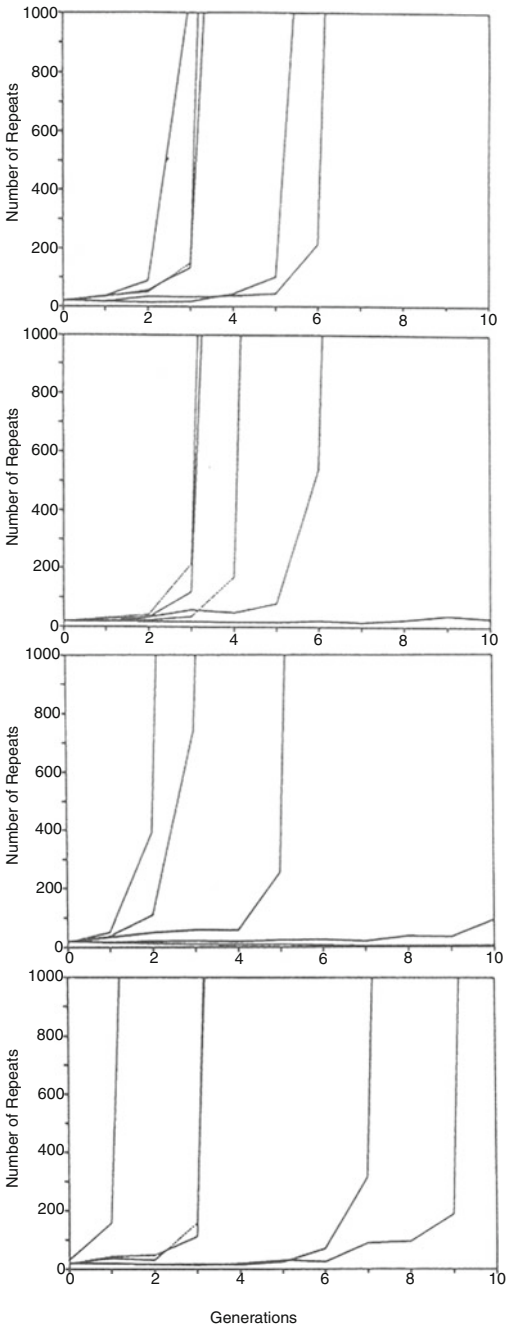


Fig. 3.9 Twenty simulation runs of the iterated Galton–Watson process with binomial thinning. Parameters are $p = 0.05$ and $u = 0.8$, i.e., a new "initiation before termination" event occurs with probability 5% and the efficiency of the resolution and reincorporation process is 80%. Each run starts from exactly 20 repeats and continues to fluctuate within narrow limits for a variable number of generations. Once the fluctuation exceeds $100 - 200$ repeats, it usually jumps to ≥ 1000 repeats. (Source: Gawel and Kimmel 1996)

3.9 Application: Galton–Watson Processes in Random Environment and Macroevolution

In evolutionary biology, it is frequently assumed that the environment of a population is fluctuating randomly (Gillespie 1996). If the dynamics of a population is described by a branching process, this means that the pgf of the number of progeny per particle varies randomly from one generation to another.

The following account follows unpublished lecture notes by V. Vatutin (personal communication): Assume that the reproduction law in a GW process is changing from generation to generation and, particles of mth generation produce offspring according to the pgf $f_m(s)$. Clearly,

$$F_n(s) = F_{n-1}[f_n(s)] = f_0(f_1(\cdots(f_n(s))\cdots))$$

is the pgf specifying the distribution law of Z_n. One important case is the randomly changing environment. Specifically, let us define a collection $\mathcal{G} = \{G_a : a \in \mathcal{A}\}$ of pgfs with $\mathcal{A}$ being some set. The reproduction law of the particles of the ith generation is taken from $\mathcal{G}$ at random according to some law

$$f_i \in \mathcal{G}, \text{ iid.}$$

Let us note that this setup implies dependent reproduction in successive generations. The above model is called the GW branching processes in random environment (GWBPRE). Let

$$\rho = \mathrm{E}[\ln f_0'(1-)].$$

The GWBPRE is said to be subcritical if $\rho < 0$, critical if $\rho = 0$, and supercritical if $\rho > 0$. For nontriviality, we assume that

$$\mathrm{Var}[\ln f_0'(1-)] > 0.$$

3.9.1 Reduced Trees for Subcritical GWBPRE

The concept of reduced process is important for the reversed-time analysis of branching processes. It involves the part of the process that contributed to individuals seen in the present time. Mathematically, we define the reduced process (backward genealogical tree) as a family

$$\{Z_{m,n},\ 0 \le m \le n\}$$

in which $Z_{m,n}$ is the number of particles at time $m \in [0, n]$ with nonempty offspring at time n.

Fleischmann and Vatutin (1999) established that for the linear-fractional case (Sect. 3.1.4) and $m > 1$, we have

$$\lim_{n\to\infty} \mathrm{P}[Z_{m,n} = k \mid Z_n > 0] = q_k(m) > 0, \ \sum_{k=1}^{\infty} q_k(m) = 1,$$

and for all $m^* > 0$ we have

$$\lim_{n\to\infty} \mathrm{P}[Z_{n-m^*,n} = k \mid Z_n > 0] = q_k^*(m^*) > 0, \ \sum_{k=1}^{\infty} q_k^*(m^*) = 1,$$

and, finally, if u_n and v_n are such that

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} v_n = \infty, \ \lim_{n\to\infty} (n - u_n - v_n) = \infty,$$

then

$$\lim_{n\to\infty} \mathrm{P}[Z_{u_n,n} = Z_{n-v_n,n} \mid Z_n > 0] = 1. \tag{3.38}$$

Let us assume that the present time is n, in the units of one generation of particles. If we observe a nonextinct process population that evolved in the past like a subcritical GWBPRE, we see that with a high probability, during the long time interval $[u_n, n - v_n]$, the process did not change state. This means that the divergence happened either very close to the present moment or very far in the past.

3.9.2 *Evolutionary Interpretation*

V. Vatutin (personal communication) noticed that (3.38) may enable a reinterpretation of conclusions based on molecular evidence of genetic divergence between humans and chimpanzees. One of the more influential recent evolutionary theories is the theory of punctuated equilibria. The theory, based on some fossil evidence, states that long periods of evolutionary stasis (invariance of species) are interspersed with bursts of speciation (appearance of new species). If the evolutionary process can be modeled using a subcritical GWBPRE, then the observed periods of evolutionary stasis preceded and followed by bursts of speciation may not necessarily reflect the unevenness of the evolutionary process itself, but follow from the properties of the reduced GWBPRE. Gillespie's (1986) more general observations concerning the evolution's "episodic clock" can be similarly reinterpreted. Gillespie (1986) has investigated the ratio R of the variance to the mean in a set of four nuclear and five mitochondrial genes in mammals, and found that R ranged from 0.16 to 35.55, which can be interpreted as periods of stasis alternating with periods of rapid substitution. To fit these data Gillespie (1986) suggested models that incorporate natural selection in a changing environment. Reduced GWBPRE might provide an alternative for these models.

3.10 Other Works and Applications

Much work has been done concerning both various generalizations of the GW process and diverse properties of the basic process. Further in this book, we will consider examples of Galton–W processes with diverse type spaces. In this section, we provide examples of a different kind.

3.10.1 Stochastic Dependence

Stochastic dependence in branching processes can be formulated in various ways. Examples include, intergeneration dependence and dependence between relatives. Both are interesting because of their applications in cell proliferation. It is known that progeny cells emerging from a division of a parent cell have lifelengths and other parameters which are correlated. A number of researchers attempted to account for these empirical observations (Axelrod and Kuczek 1989; Brooks et al. 1980; Rigney 1981; Hejblum et al. 1988; Kuczek and Axelrod 1986; Sennerstam and Strömberg 1996; Staudte et al. 1984; Staudte et al. 1997; Webb 1989).

Generation dependence (Fearn 1972) has a different meaning for the GW process in which the generations are synchronized, and in the time-continuous age-dependent processes, in which the generations overlap (Chaps. 4 and 5). One way of capturing dependence between relatives is to consider the individual together with his/her relatives (siblings, cousins, etc.) as a single superindividual. This can be carried out using the framework of general processes (Olofsson 1996). In the framework of estimation, a convenient manner of expressing such "local" dependencies is the bifurcating autoregression (Sect. 5.5.4).

3.10.2 Process State Dependence

All GW processes, including these for which the progeny distributions depend on the state of the process, are Markov. However, there is no simple relationship linking the type of dependence with the properties of the resulting Markov chain. Therefore, the study of such processes proceeds by way of special cases, deemed important usually for extramathematical reasons. An early reference is Lipow (1975).

A series of papers by Klebaner consider limit properties of processes with progeny distributions depending on the process state (i.e., usually the number of particles at a given time). Klebaner (1996) is a short review of size- and density-dependent processes. Klebaner (1988, 1990) and Klebaner and Cohn (1986) consider applications in demography and genetics. Another interesting paper (Klebaner and Zeitouni 1994) considers the problem of "cycle slip", i.e., the conditions that a randomly perturbed

deterministic system has to satisfy to escape the basin of attraction of the deterministic part. A more recent paper by Klebaner (2010) concerns approximations of state-dependent branching process.

Another application is presented by Jagers (1995) and Haccou et al. (2005) who used the coupling method to analyze state-dependent processes describing proliferation of biological cells.

3.10.3 Bisexual Galton–Watson Process

The bisexual generalization of the GW process is not straightforward to consider, because it involves a process of pair formation. One way to proceed is to assume that only females bear progeny, of both genders, and to define a mating function which provides for each unpaired female the probability of forming a pair and mating with an available male. These functions may be consistent with monogamy or monoandry or they may mimic the mating patterns of insects, etc. The mating process destroys the branching property and the resulting stochastic process is not strictly speaking a branching process. One of the papers on the limit properties of such processes is González and Molina (1996). A related paper (González et al. 2010), from the same group at the University of Extremadura at Badajoz, concerns statistical inference for Y-linked gene branching models using the expectation-maximization method. An exhaustive review of older literature is provided in the thesis by Falahati (1999) and of more recent literature in Molina et al. (2010).

3.10.4 Age of the Process

Estimation of the age of the branching process based on data concerning extant individuals, their number, types, etc., gained importance because of applications in genetics and molecular evolution. Evolution of chromosomes containing disease genes can be represented as a branching process with Poisson distribution of progeny, if the disease subpopulation is a small subset of a larger population evolving according to the Fisher–Wright model. A model of this type was considered by Kaplan et al. (1995) and used to obtain simulation-based likelihood estimates of location and age of disease genes. A number of refinements of this method can be found in the unpublished doctoral thesis of Pankratz (1998), where further references also are provided.

Another type of application is finding the age of the most recent common ancestor of a population characterized by its genetic makeup, under the assumption that its demography followed a branching process. An example related to evolution of modern humans, using a time-continuous branching process, is described in detail in Sect. 8.3.

An early paper concerning estimation of the age of a GW branching process is Stigler (1970). The author uses the linear-fractional case, in which an estimator can be explicitly derived, and then generalizes the results to the case of the general GW process. This paper was followed by a number of other publications, including Tavaré (1980) and Koteeswaran (1989).

3.10.5 Family Trees and Subtrees

A somewhat related subject is the probability that the family tree of the process contains an infinite N-nary subtree, i.e., a tree with exactly N progeny of each individual. Pakes and Dekking (1991) demonstrated that this probability is the largest root in the interval $[0, 1]$ of the equation

$$1 - t = G_N(1 - t),$$

where

$$G_N(s) = \sum_{j=0}^{N-1} (1 - s)^j f^{(j)}(s)/j!,$$

and $f^{(j)}(s)$ is the offspring distribution of the process. Further results concerning the maximum height of the N-nary subtree are provided in the same paper.

3.10.6 Model of Next Generation Sequencing

Heinrich et al. (2012) model the DNA next generation sequencing (NGS) procedure as a branching process and derive a mathematical framework for the expected distribution of alleles at heterozygous loci before sequencing. This description takes into account the process of "proliferation" of copies of oligonucleotides. The resulting GW branching process has a larger variance than that expected from binomial or Poisson counting. Theoretical results are confirmed by analyzing technical replicates of human exome data, computing the variance of allele frequencies at heterozygous loci. Due to this high variance, mutation callers relying on binomial distributed priors are less sensitive for heterozygous variants that deviate strongly from the expected mean frequency. The results also indicate that error rates can be reduced to a greater degree by technical replicates than by increasing sequencing depth (since technical replicates are stochastically independent, while the opposite is the case for reads obtained from NGS).

3.11 Problems

1. Below are given several examples of pgf of a GW process. For each of them, find $\mathrm{E}(Z_1) \equiv m$ and $\mathrm{Var}(Z_1) \equiv \sigma^2$. Assume that the GW process describes a cell population with discrete generations. Characterize the model described by each pgf. *Example:* If $f(s) = (ps + q)^2$ then each of the two daughter cells, independently, survives with probability p and dies with probability q.
   - $f(s) = ps^2 + qs$,
   - $f(s) = ps^2 + q$,
   - $f(s) = ps^2 + qs + r$.
2. Assume that the pgf of the GW process is the fractional linear function. Using induction, prove the form of $f_n(s)$ in the case $m = 1$.
3. Assume the fractional linear case. Treating the GW process as a Markov chain, check that the state $\{Z_n = k\}$ is *transient* if $k \neq 0$ and recurrent if $k = 0$. *Hint:* Use the closed form of $f_n(s)$ and base the assertion on the condition of divergence of $\sum_{n\geq 0} \Pr\{Z_n = k\}$.
4. Suppose that a GW process with the pgf $f(s)$ is started not by a single particle, but by a random number of particles (with pgf $g(s)$). Find $f_n(s)$.
5. *Continued.* Assume $f(s)$ the linear-fractional function with $m < 1$ and $g(s) = \frac{(q-1)s}{q-s}$, where $q = f(q)$. Define $\bar{f}_{(n)}(s) = \frac{f_n(s)-f_n(0)}{1-f_n(0)}$ the conditional pgf of Z_n provided $Z_n > 0$. Prove, using induction, that $\bar{f}_{(n)}(s) \equiv g(s)$.
6. Distribution with pgf $g(s)$ having properties as above, is called *a quasistationary distribution* of the GW process. What makes it different compared to the stationary distribution of a Markov chain?
7. *Galton–Watson process in varying environment*. Suppose that the nth generation of particles has the progeny distribution $\{p_k^{[n]},\ k \geq 0\}$ with pgf $f^{[n]}(s)$. Define the process in the terms of a Markov chain and derive the forward equation as it was done for the ordinary process. What is $f_n(s)$ now?
8. *Integrated GW process*. Consider the process $\{Y_n\}$, where $Y_n = \sum_{i=0}^{n} Z_i$. Demonstrate that the pgf of Y_n, denoted $F_n(s)$, satisfies

$$F_{n+1}(s) = sf[F_n(s)].$$

9. *Continued.* Demonstrate that if $m < 1$, then the limit $\lim_{n\to\infty} F_n(s) = F(s)$ exists and satisfies the following functional equation:

$$F(s) = sf[F(s)].$$

   Hint. Show that $|F_{n+1}(s) - F_n(s)| \leq m|F_n(s) - F_{n-1}(s)|$ if $s \in [0, 1]$. $F(s)$ is the pgf of the total number of particles produced in the process and in the subcritical case it is a proper random variable, i.e., $F(1) = 1$.
10. *Continued.* Assume the linear-fractional case and calculate $F(s)$ by solving the functional equation above. Does $F(s)$ correspond to any standard discrete distribution?
11. *Quasistationary distribution.* Suppose that a subcritical GW process with the pgf $f(s)$ is started not by a single particle, but by a random number of particles

having pgf $\mathcal{B}(s)$, defined in the Yaglom theorem. Prove that this distribution is a stationary distribution of the GW process. *Hint.* Use the functional equation defining $\mathcal{B}(s)$ and the property that $\mathcal{B}(0) = 0$.

12. Assume the linear-fractional case and $m > 1$. Calculate the Laplace transform of $W_n = \frac{Z_n}{m^n}$ and find its limit as $n \to \infty$. What is the distribution of W?
13. *Consider the following mechanism of gene amplification:*
    - Each of the double minute chromosomes present in the newborn daughter cell survives wp p. If it does survive, then during *replication* each next copy of this particular double minute chromosome is produced wp p.
    - During *segregation* each copy is assigned to given daughter cell wp $\frac{1}{2}$.

    Consider a random lineage of cells in the population. If in the zeroth generation there exists only a single cell with a single double minute chromosome, then $\{Z_n, n \geq 0\}$ the sequence of number of copies of the double minute in the cell of nth generation, forms a GW process with the progeny pgf $f(s)$. Find $f(s)$. *Hint.* Use the expression for the pgf of the sum of random number of iid rv's.
14. *Continued.* Using the properties of the linear-fractional pgf's, assuming the subcritical process, find the pgf $\mathcal{B}(s)$ of the limit distribution of the number of double minute chromosomes per cell in the cells of the resistant clone. Suppose the mean number of double minute chromosomes per resistant cell is equal to 20 and that double minutes have been counted in 50 cells. Find the maximum likelihood estimate of p and an approximate 95% confidence interval for this estimate.
15. Consider a population of particles with lifelengths equal to 1, proliferating by binary fission, with each of the two progeny surviving independently with probability p.
    a) Find the probability of eventual extinction

    $$q = \Pr\{\#\text{ particles } = 0 \text{ at some time } n\},$$

    for a population started by a single ancestor particle, as the function of p (i.e., $q = q(p)$), for $p \in [0, 1]$.
    b) Find the probability that at time $n = 3$, there will be 4 or less particles in the process.
    c) An ad hoc way to increase the probability of non-extinction of the process is to start at time 0 from a collection of N ancestor particles, instead of 1. Find the probability $q = q(p, N)$ of eventual extinction of such process. For $p = 3/4$, what should be N equal to so that $1 - q(p, N)$ exceed 0.999?
16. Consider a GW process Z_n with progeny pgf $h(s)$, started by a random number Y of ancestors [where $Y \sim g(s)$]. Find
    a) $\mathrm{E}\,(Z_n | Z_0 = Y)$.
    b) $\mathrm{Var}(Z_n | Z_0 = Y)$.
    c) $\Pr\{Z_n = 0, \text{ some } n \,| Z_0 = Y\}$.

Chapter 4
The Age-Dependent Process: Markov Case

This chapter is devoted to the use of time-continuous branching process with exponential life-time distributions. This process also has the Markov property and is closely related to the Galton–Watson process. The exponential distribution to model lifetimes of particles is not well motivated by any biological assumptions. Indeed, the exponential distribution admits lifetimes which are arbitrarily close to 0, while it is known that life cycles of organisms and cells have lower bounds of durations, which are greater than 0. The advantage of using the exponential distribution is that it leads, in many cases, to computable expressions. These expressions allow one to deduce properties which can then be conjectured for more general models.

4.1 Differential Equation for the pgf and its Elementary Properties

4.1.1 Definition of the Process

The process can be described as follows. A single ancestor particle is born at $t = 0$. It lives for time τ, which is exponentially distributed with parameter λ. At the moment of death, the particle produces a random number of progeny according to a probability distribution with pgf $f(s)$. Each of the first-generation progeny behaves, independently of each other, in the same way as the initial particle. It lives for an exponentially distributed time and produces a random number of progeny. Progeny of each of the subsequent generations behave in the same way. If we denote $Z(t)$ the particle count at time t, then we obtain a stochastic process $\{Z(t),\ t \geq 0\}$.

The probability-generating function $F(s,t)$ of $Z(t)$ satisfies an ordinary differential equation which is easiest to derive based on the Markov nature of the process. Indeed, let us consider the process at a given time t. Any of the particles existing at this time, whatever its age is, has a remaining lifetime being distributed exponentially with parameter λ. This follows from the lack of memory of the exponential distribution. Therefore, each of the particles starts, independently, a subprocess, identically distributed with the entire process (Fig. 4.1). Consequently, at any time $t + \Delta t$, the

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_4

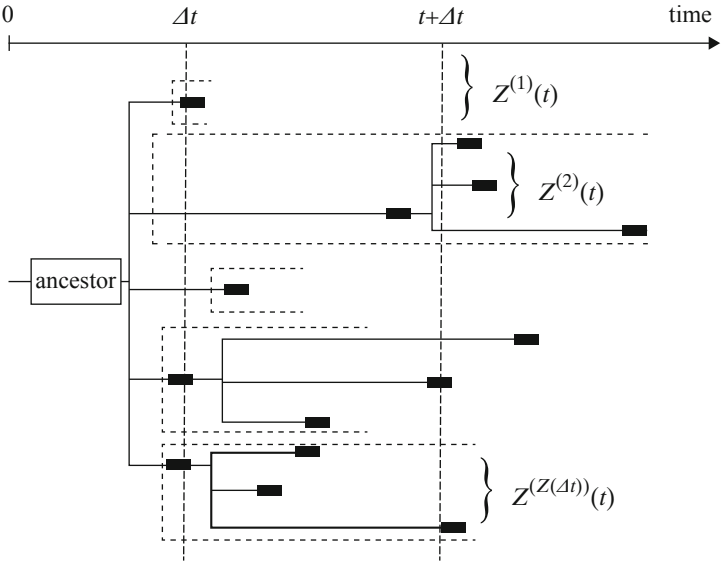


Fig. 4.1 Derivation of the backward equation for the Markov time-continuous branching process

number of particles in the process is equal to the sum of the number of particles in all independent identically distributed (iid) subprocesses started by particles existing at time Δt. Each of these subprocesses is of age t. In mathematical terms,

$$Z(t+\Delta t) = \sum_{i=1}^{Z(\Delta t)} Z^{(i)}(t), \tag{4.1}$$

where superscript (i) identifies the i-th iid subprocess. So, according to the pgf theorem (Theorem 1.1), we have the following pgf identity:

$$F(s, t+\Delta t) = F[F(s,t), \Delta t]. \tag{4.2}$$

We subtract $F(s,t)$ from both sides and, remembering that the process is started by a single particle, i.e., $F(s,0) = s$, we can write the result in the following form:

$$F(s, t+\Delta t) - F(s,t) = F[F(s,t), \Delta t] - F[F(s,t), 0]. \tag{4.3}$$

If Δt is small, then with a probability close to 1, the process consists only either of the ancestor or of its first-generation progeny. In the terms of the process pgf,

$$F(s, \Delta t) = s\mathrm{e}^{-\lambda\Delta t} + f(s)(1 - \mathrm{e}^{-\lambda\Delta t}) + o(\Delta t), \tag{4.4}$$

or

$$F(s, \Delta t) - F(s,0) = [-s + f(s)](1 - \mathrm{e}^{-\lambda\Delta t}) + o(\Delta t). \tag{4.5}$$

Substituting (4.5) into (4.3) and dividing by Δt we obtain

$$\frac{F(s,t+\Delta t)-F(s,t)}{\Delta t}=\frac{\{-F(s,t)+f[F(s,t)]\}(1-\mathrm{e}^{-\lambda\Delta t})+o(\Delta t)}{\Delta t}.$$

By letting $\Delta t \to 0$, this leads to the following differential equation:

$$\mathrm{d}F(s,t)/\mathrm{d}t=-\lambda\{F(s,t)-f[F(s,t)]\}. \tag{4.6}$$

Equation (4.6), with the initial condition $F(s,0)=s$, has a unique pgf solution if conditions are satisfied, which guarantee that the process does not explode, i.e., at each time $t>0$, the number of particles is finite wp 1 or $\lim_{s\uparrow 1} F(s;t)=1$. For this, it is sufficient that the expected number of progeny per particle $m=f'(1-)$ is finite (Athreya and Ney 2004).

In particular, expression (4.2) demonstrates that for any time increment Δt, we have

$$F(s,i\Delta t)=f_{\Delta t}^{(i)}(s),$$

where $f_{\Delta t}^{(i)}(s)$ is the ith iterate of $F(s,\Delta t)$. Therefore, $\{Z(i\Delta t,\omega),\ i=0,1,\dots\}$ is a Galton–Watson process with progeny pgf $f_{\Delta t}(s)$. Of course, $f_{\Delta t}(s)$ has properties very different from those of $f(s)$. In particular, even if $f(s)$ admits only a finite number of progeny, $f_{\Delta t}^{(i)}(s)$ always has an infinitely long right tail.

4.1.2 *Probability of Extinction and Moments*

The Markov branching process is called

- *Subcritical*, if $m<1$
- *Critical*, if $m=1$
- *Supercritical*, if $m>1$

Let q be defined as for the Galton–Watson process, i.e., as the smallest root of the equation $f(s)=s$, $s\in[0,1]$. The extinction probability is, again, equal to q.

Theorem 4.1 *Suppose $m<\infty$. If $F(s;t)$ is the pgf solution of Eq. (4.6), then $P(t)\equiv F(0;t)\to q$ as $t\to\infty$.*

The extinction probability result is the same as for the Galton–Watson process. The expressions for the moments are almost as simple as they are for the Galton–Watson process.

Let us define the *k-th factorial moment of* $Z(t)$, $m_k(t)=\mathrm{E}\{Z(t)[Z(t)-1]\cdots[Z(t)-k+1]\}$. The differential equations for the factorial moments of the process are formally derived by differentiating Eq. (4.6) with respect to s and letting $s\uparrow 1$. For example, the expected value $m_1(t)$ satisfies,

$$\frac{\mathrm{d}m_1(t)}{\mathrm{d}t}=\lambda(m-1)m_1(t),\quad m_1(0)=1.$$

These equations can be solved explicitly. We obtain the following expressions for the expectation and variance of $Z(t)$:

$$\mathrm{E}[Z(t)] = \mathrm{e}^{at}, \tag{4.7}$$

$$\mathrm{Var}[Z(t)] = \begin{cases} \frac{f''(1-)-f'(1-)+1}{f'(1-)-1}\mathrm{e}^{at}(\mathrm{e}^{at}-1), & a \neq 0 \\ f''(1-)\lambda t, & a = 0, \end{cases} \tag{4.8}$$

where $a = \lambda(f'(1-)-1)$ is the *Malthusian parameter of population growth.*

4.2 Application: Clonal Resistance Theory of Cancer Cells

The aim of cancer chemotherapy is to achieve remission, i.e., disappearance of clinically detectable cancers and then to prevent relapse, i.e., the regrowth of cancer. In many cases, the failure of chemotherapy is associated with the growth of cells resistant to further treatment with the same drug. There are two conceivable modes of drug resistance: Resistant cells might exist in tumors before treatment and might be selected for during treatment. Alternatively, they might be induced by treatment.

Drug resistance was extensively studied in bacteria (see Sect. 6.1 and also a review paper by Levy 1998), and the resulting ideas have been applied to understand drug resistance in cancer cells. One possible hypothesis is that mutations from sensitivity to resistance are rare, irreversible events, that spontaneously occur in the absence of the selecting drug. Moreover, mutation in resistance to a drug is a single event, and it arises independently of resistance to another drug. Although simplistic, this model is useful in understanding the initiation and growth of drug-resistant cancer cells. Also, it might help design new protocols of cancer chemotherapy.

We explore the branching process approach to a theory of resistance, which has become influential in the cancer research community. It was originally developed by Coldman and Goldie (Goldie and Coldman 1979; Coldman and Goldie 1985). We will re-derive some of the original results, using Markov time-continuous branching processes. This approach seems more rigorous.

The assumptions of the theory are as follows (Fig. 4.2):

1. The cancer cell population is initiated by a single cell which is sensitive to the cytotoxic (chemotherapeutic) agent. The population proliferates without losses.
2. Interdivision time of cells is a random variable with a given distribution.
3. At each division, with given probability, a single progeny cell mutates and becomes resistant to the cytotoxic agent.
4. Mutations are irreversible.

We wish to compute the probability that when the tumor is discovered, it does not contain any resistant cells. Only in such a situation, is the use of a cytotoxic agent effective. If even a small subpopulation of resistant cells exists, the cancer cell population will eventually reemerge despite the therapy.

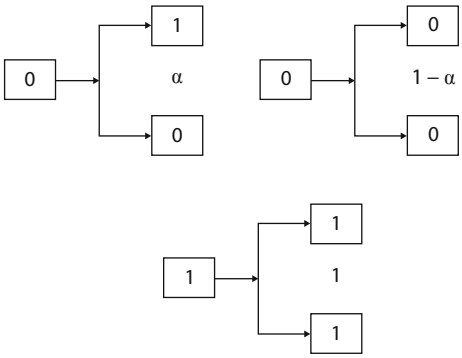


Fig. 4.2 Schematic representation of the branching process of clonal resistance, in the single-mutation case

4.2.1 Single-Mutation Case

The Branching Process Model

We translate the hypotheses of clonal resistance into the language of branching processes.

1. In the process, there exist two types of particles, labeled 0 (sensitive) and 1 (resistant).
2. The process is initiated by a single type 0 particle.
3. The life spans of particles are independent random variables, distributed exponentially with parameter λ.
4. Each particle, at death, divides into exactly two progeny particles:
   - 0-particle produces either two 0-particles, wp $1 - \alpha$, or one 0− and one 1-particle, wp α.
   - 1-particle produces two 1-particles.

Thus, we have a *two-type* time-continuous Markov branching process.

Let us introduce the following notations, which are required since we consider two types of particles:

- $F_0(s_0, s_1; t)$ is the joint probability-generating function (see Appendix A) of the numbers of cells of both types, present at time t, *in the process initiated at time* 0 *by a type 0 cell.*
- $F_1(s_1; t)$ *is the pgf of the numbers of cells of type* 1*, present at time* t*, in the process initiated at time* 0 *by a type 1 cell.*

Frequently, we will be writing $F_i(s; t)$ and even $F_i(t)$ or $F_i(s)$ or F_i.

In the general case of the process with k types of particles, we denote $f_i(s) = f_i(s_1, \ldots, s_k)$, the joint pgf of the number of progeny of all k types begotten by an i-type particle. The lifetime of an i-type particle is exponentially distributed with parameter λ_i. Denoting by $F_i(s; t)$, the joint pgf of the number of particles of all

types in a process started by an ancestor of type i, we write the system of ordinary differential equations

$$\mathrm{d}F(s;t)/\mathrm{d}t = -\lambda \cdot \{F(s;t) - f[F(s;t)]\}, \tag{4.9}$$

in which F, f , and λ are vectors and operator "$\cdot$" is a componentwise product of two vectors. The initial condition is $F(s;0) = s$.

In our application, based on hypothesis 4, $f_0(s) = (1-\alpha)s_0^2 + \alpha s_0 s_1$, $f_1(s) = s_1^2$, and $\lambda_0 = \lambda_1 = \lambda$. In consequence,

$$\frac{\mathrm{d}F_0}{\mathrm{d}t} = -\lambda F_0 + \lambda[(1-\alpha)F_0^2 + \alpha F_0 F_1], \tag{4.10}$$

$$\frac{\mathrm{d}F_1}{\mathrm{d}t} = -\lambda F_1 + \lambda F_1^2. \tag{4.11}$$

Solutions

Finding explicit solutions for cell proliferation models of the type (4.10), (4.11), frequently lead to differential equations with right-hand sides quadratic in the unknown function (so-called Riccatti-type equations). The reason is that in such models the pgf of the number of progeny is a second-order polynomial, which reflects the binary fission mode of proliferation of living cells. The following result can be verified by direct substitution: Uniqueness follow by the usual regularity conditions.

Theorem 4.2 *The solution of the differential equation*

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = f(t)F(t) + hF(t)^2, \tag{4.12}$$

where $f \in C[0,\infty)$, with initial condition $F(0)$, is a uniquely defined function $F \in C^1[0,\infty)$

$$F(t) = \frac{F(0)\mathrm{e}^{\int_0^t f(u)\mathrm{d}u}}{1 - hF(0)\int_0^t \mathrm{e}^{\int_0^u f(v)\mathrm{d}v}\mathrm{d}u}. \tag{4.13}$$

We will solve the system (4.10), (4.11). First, separation of variables, or Eq. (4.13) is applied to Eq. (4.11) and it yields

$$F_1(s;t) = \frac{s_1}{s_1 + (1-s_1)\mathrm{e}^{\lambda t}}. \tag{4.14}$$

Substituting (4.14) into Eq. (4.10) and employing Theorem 4.2, we obtain

$$F_0(s;t) = \frac{s_0\mathrm{e}^{-\lambda t}[\mathrm{e}^{-\lambda t}s_1 + (1-s_1)]^{-\alpha}}{1 + s_0\{[\mathrm{e}^{-\lambda t}s_1 + (1-s_1)]^{1-\alpha} - 1\}s_1^{-1}}. \tag{4.15}$$

Differentiating $F_0(s;t)$ with respect to s_0 and s_1, we obtain the expressions for the expected counts of the sensitive and resistant cells

$$M_0(t) = \frac{\partial F(1,1;t)}{\partial s_0} = \mathrm{e}^{\lambda(1-\alpha)t}, \quad t \geq 0,$$

$$M_1(t) = \frac{\partial F(1,1;t)}{\partial s_1} = \mathrm{e}^{\lambda t} - \mathrm{e}^{\lambda(1-\alpha)t}, \quad t \geq 0.$$

The conclusion is that in absence of intervention, the resistant cells eventually outgrow the sensitive ones. The probability of no-resistant cells at time t is also easy to obtain

$$P(t) = \lim_{s_0 \uparrow 1} \lim_{s_1 \downarrow 0} F_0(s;t) = \frac{1}{(1-\alpha) + \alpha \mathrm{e}^{\lambda t}} = \frac{1}{(1-\alpha) + \alpha[M_0(t) + M_1(t)]}. \tag{4.16}$$

Conclusions

Based on Eq. (4.16), the following observations can be made:

- The probability that there are no resistant cells at time t is inversely related to the total number of cells.
- For different mutation rates α, if α's are small, the plots of $P(t)$ are approximately shifted, with respect to each other, along the t axis.
- The time interval in which the resistant clone is likely to emerge, i.e., in which $P(t)$ falls from near 1 to near 0, for example, from 0.95 to 0.05, constitutes a relatively short "window" (Fig. 4.3). Therefore, the therapy should be prompt and radical to decrease cell number and probability $(1 - P(t))$ of emerging resistance.

An Alternative Model

An alternative variant of the model presented above assumes that each of the progeny cells may mutate independently with probability α, as depicted in Fig. 4.4. The equations of the process assume now the form

$$\frac{\mathrm{d}F_0}{\mathrm{d}t} = -\lambda F_0 + \lambda[(1-\alpha)F_0 + \alpha F_1]^2, \tag{4.17}$$

$$\frac{\mathrm{d}F_1}{\mathrm{d}t} = -\lambda F_1 + \lambda F_1^2. \tag{4.18}$$

They are of a more general Riccatti form, not admitting a closed-form solution. However, it is still possible to obtain $P(t)$. Let us note that $F_1(1,0;t) \equiv 0$, i.e., the probability of no-resistant cells in the subprocess initiated by a resistant cell is

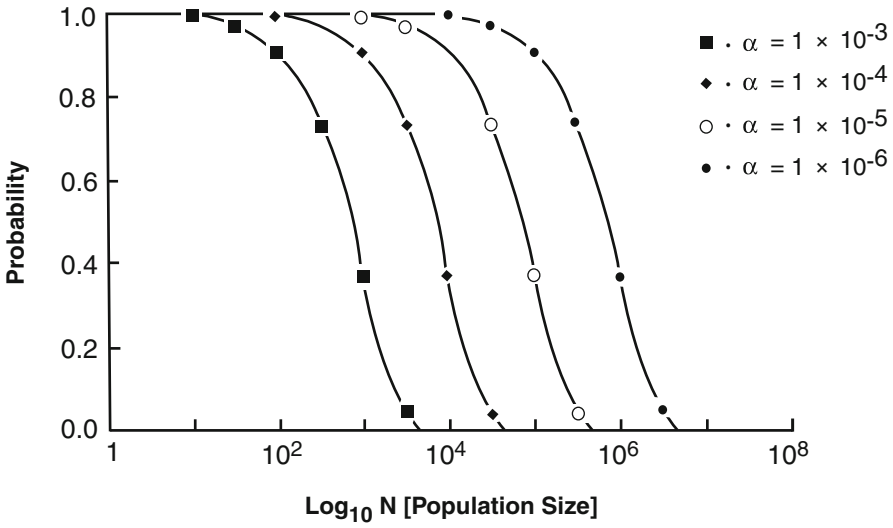


Fig. 4.3 Probability $P(t)$ of no-resistant cells, depending on mutation rate and tumor size $N(t) = \exp(\lambda t)$, in the single-mutation model. (Source: Coldman and Goldie 1985)

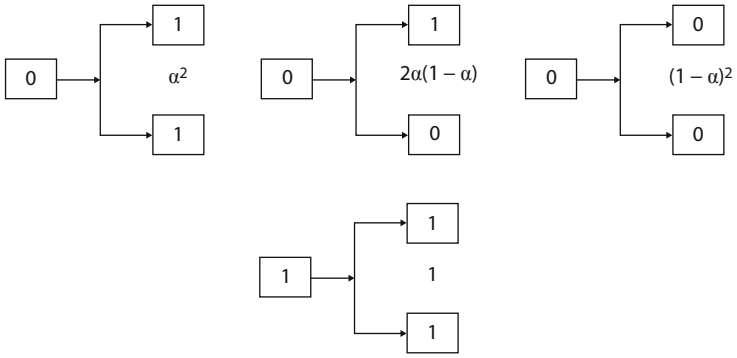


Fig. 4.4 Schematic representation of the *alternative* branching process of clonal resistance, in the single-mutation case

equal to 0. Therefore, letting $s_0 \uparrow 1$ and $s_1 \downarrow 0$ in Eq. (4.17) yields the following differential equation for $P(t)$:

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = -\lambda P(t) + \lambda(1-\alpha)^2 P(t)^2; \quad P(0) = 1, \tag{4.19}$$

the solution of which is

$$P(t) = \frac{1}{[1-(1-\alpha)^2]\mathrm{e}^{\lambda t} + (1-\alpha)^2}; \quad t \geq 0. \tag{4.20}$$

If α is small and consequently α^2 is a second-order small, then the new $P(t)$ is approximately equal to that in (4.16), with α replaced by 2α.

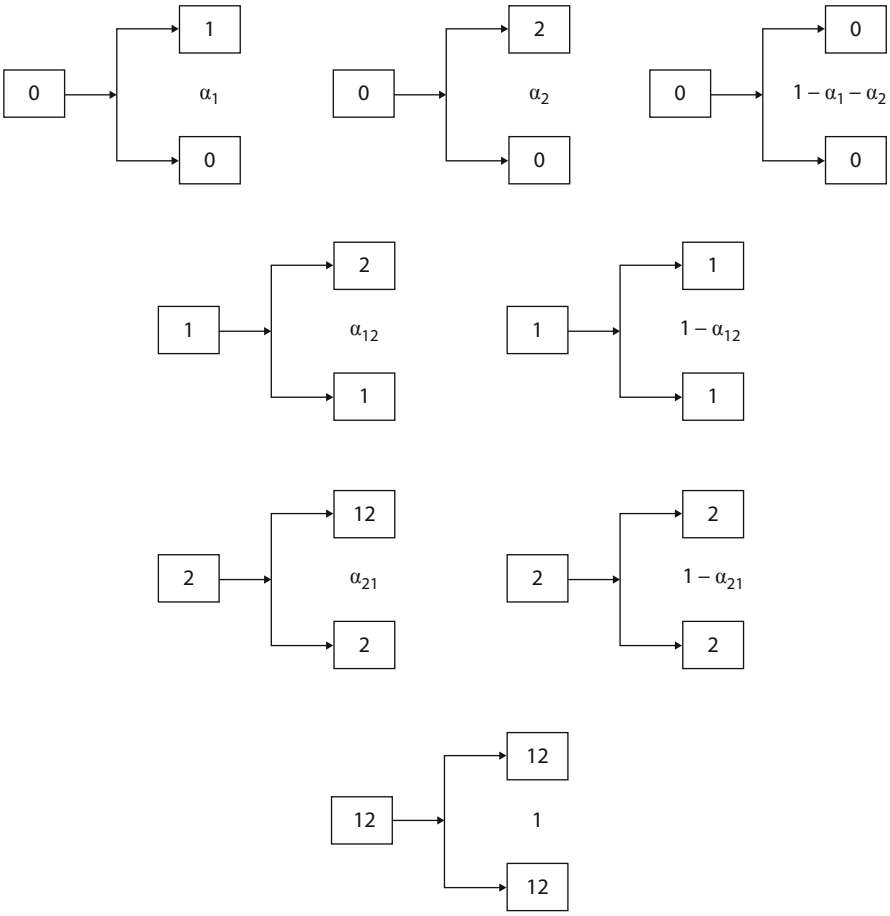


Fig. 4.5 Schematic representation of the branching process of clonal resistance, in the two-mutations case

4.2.2 Two-Mutations Case

The aim of the two-mutation model is to address the problem of the so-called cross resistance, i.e., resistance to more than one cancer-cell-killing agent. Cross resistance is important for cancer chemotherapy, since protocols including more than one agent are frequently used in therapy.

We will specify the hypotheses of our model (Fig. 4.5).

- The population of cells proliferates by binary fission starting from a single cell. The lifespans of all the cells are independent, exponentially distributed, random variables with parameter λ.
- The founder cell of the population is *sensitive* to chemotherapy.
- A sensitive cell divides into either two sensitive cells or one sensitive cell and the other cell *resistant to drug 1*, or it divides into one sensitive cell and the other cell

resistant to drug 2. These events occur with respective probabilities $1-\alpha_1-\alpha_2$, α_1, and α_2.
- A cell resistant to drug 1 divides into either two cells resistant to drug 1 or one cell resistant to drug 1 and the other *resistant to drugs 1 and 2*. These events occur with respective probabilities $1-\alpha_{12}$ and α_{12}.
- A cell resistant to drug 2 divides into either two cells resistant to drug 2 or one cell resistant to drug 2 and the other *resistant to drugs 1 and 2*. These events occur with respective probabilities $1-\alpha_{21}$ and α_{21}.
- A cell resistant to drugs 1 and 2 divides into two cells resistant to drugs 1 and 2.

We will name the sensitive cells as *type 0*, cells resistant to drug 1 as *type 1*, cells resistant to drug 2 as *type 2*, and cells resistant to drugs 1 and 2 as *type 12*, respectively.

The rules specified above define a 4-type time-continuous Markov branching process. The mathematical description of this process is based on the observation that it can be decomposed into unions of subprocesses generated by progeny cells of different types. There are four types of such subprocesses, generated by cells of type 0, 1, 2, and 12, respectively. Biologically, they can be identified with clones of different cells. Let us introduce the following notations:

- $F_0(s_0,s_1,s_2,s_{12};t)$ is the joint pgf of the numbers of cells of all types, present at time t in the process initiated by a type 0 cell. This particular subprocess is identical, in distribution, with the entire process.
- $F_1(s_0,s_1,s_2,s_{12};t)$ is the joint pgf of the numbers of cells of all types, present at time t in the process initiated by a type 1 cell.
- $F_2(s_0,s_1,s_2,s_{12};t)$ is the joint pgf of the numbers of cells of all types, present at time t in the process initiated by a type 2 cell.
- $F_{12}(s_0,s_1,s_2,s_{12};t)$ is the joint pgf of the numbers of cells of all types, present at time t in the process initiated by a type 12 cell.

We obtain the following system of ordinary differential equations for the probability-generating functions F_0, F_1, F_2, and F_{12}:

$$\frac{\mathrm{d}F_0}{\mathrm{d}t} = -\lambda F_0 + \lambda[(1-\alpha_1-\alpha_2)F_0^2 + \alpha_1 F_0 F_1 + \alpha_2 F_0 F_2], \tag{4.21}$$

$$\frac{\mathrm{d}F_1}{\mathrm{d}t} = -\lambda F_1 + \lambda[(1-\alpha_{12})F_1^2 + \alpha_{12} F_1 F_{12}], \tag{4.22}$$

$$\frac{\mathrm{d}F_2}{\mathrm{d}t} = -\lambda F_2 + \lambda[(1-\alpha_{21})F_2^2 + \alpha_{21} F_2 F_{12}], \tag{4.23}$$

$$\frac{\mathrm{d}F_{12}}{\mathrm{d}t} = -\lambda F_{12} + \lambda F_{12}^2. \tag{4.24}$$

The initial conditions are $F_i(s;0) = s_i$, $i = 0,1,2,12$, where $s = (s_0,s_1,s_2,s_{12})$.

It is a little surprising that there exists a semi-explicit solution of this problem. Equation (4.24) can be solved by separation of variables. It yields:

$$F_{12}(s;t) = \frac{1}{1-(1-s_{12}^{-1})\mathrm{e}^{\lambda t}}. \tag{4.25}$$

Substituting expression (4.25) into Eqs. (4.22) and (4.23) and solving the resulting differential equations by separation of variables or application of Theorem 4.2, yields, respectively:

$$F_1(s;t) = \frac{\mathrm{e}^{-\lambda t}[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{-\alpha_{12}}}{s_1^{-1}+\{[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{1-\alpha 12}-1\}s_{12}^{-1}}, \tag{4.26}$$

$$F_2(s;t) = \frac{\mathrm{e}^{-\lambda t}[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{-\alpha_{21}}}{s_1^{-1}+\{[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{1-\alpha 21}-1\}s_{12}^{-1}}. \tag{4.27}$$

Following the substitution of Eqs. (4.25)–(4.27), Eq. (4.21) assumes the form which is solvable using Theorem 4.2. Accordingly, we calculate

$$\begin{aligned}\mathrm{e}^{\int_0^t f(u)\mathrm{d}u} &= \mathrm{e}^{-\lambda t}\{1+\{[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{1-\alpha_{12}}-1\}s_1 s_{12}^{-1}\}^{\frac{\alpha_1}{1-\alpha_{12}}} \\ &\quad \{1+\{[\mathrm{e}^{-\lambda t}s_{12}+(1-s_{12})]^{1-\alpha_{21}}-1\}s_2 s_{12}^{-1}\}^{\frac{\alpha_2}{1-\alpha_{21}}}.\end{aligned} \tag{4.28}$$

Unfortunately, $\int_0^t \mathrm{e}^{\int_0^u f(v)\mathrm{d}v}\mathrm{d}u$ cannot be obtained in a closed form. However, we are mainly interested in the probability that no doubly resistant cells emerge before t in the subprocess initiated by a sensitive cell,

$$P_{12}(t) = \mathrm{P}\{N_{12}(t)=0\} = F_0(1,1,1,0;t). \tag{4.29}$$

In this special case, expression (4.28) is reduced to

$$\begin{aligned}\mathrm{e}^{\int_0^t f(u)\mathrm{d}u} &= \mathrm{e}^{-\lambda t}[\alpha_{12}+(1-\alpha_{12})\mathrm{e}^{-\lambda t}]^{\frac{\alpha_1}{1-\alpha_{12}}} \\ &\quad [\alpha_{21}+(1-\alpha_{21})\mathrm{e}^{-\lambda t}]^{\frac{\alpha_2}{1-\alpha_{21}}}.\end{aligned} \tag{4.30}$$

The closed form solution is still not available although numerical quadrature is straightforward. However, there exists a special case of interest in which the closed form solution is available.

- Suppose that all the mutation probabilities are equal, i.e., $\alpha_1 = \alpha_2 = \alpha_{12} = \alpha_{21} = \alpha$.

In this case,

$$P_{12}(t) = \frac{\mathrm{e}^{-\lambda t}[\alpha+(1-\alpha)\mathrm{e}^{-\lambda t}]^{-\frac{2\alpha}{1-\alpha}}}{1-\frac{1-2\alpha}{1-3\alpha}\{1-[\alpha+(1-\alpha)\mathrm{e}^{-\lambda t}]^{\frac{1-3\alpha}{1-\alpha}}\}}. \tag{4.31}$$

Conclusions

Based on the model, the following observations can be made:

- For different mutation rates α, with α small, the plots of $P(t)$ are merely shifted.
- The time interval in which cross resistance is likely to emerge, i.e., in which $P_{12}(t)$ falls from near 1 to near 0, for example, from 0.95 to 0.05, constitutes a relatively short "window," similar to that in Fig. 4.3.
- It can be proved, similarly as in the one-mutation model, that the average number of cells resistant to any of the agents separately increases exponentially. Suppose that we have, at our disposal, agents 1 and 2 and that we can use them according to any time schedule, provided they are not used simultaneously. Since, in practice, only periodic chemotherapy protocols are administered, the question is, should the two drugs be alternated frequently or infrequently? The probability of double resistance emerging from cells resistant to agent 1 strongly depends on the total number of these cells. Therefore, while using agent 1, the number of cross resistant cells should be kept in check. This is more difficult if agent 1 is used for a long period without a break. The reason is that the cells resistant to agent 1 grow to large numbers, increasing the probability of cross resistance.
- Summarizing, the two agents should be alternated as frequently as possible. This is the conclusion of Goldie et al. (1982).

The original analysis of Goldie et al. (1982), replicated in this section, made use of the simplifying assumption that the two agents were equivalent in their cell-killing efficiency, i.e., $\alpha_1 = \alpha_2 = \alpha_{12} = \alpha_{21} = \alpha$. Day (1986a) confirmed the results of Goldie et al. (1982), and extended their analysis by relaxing the symmetry assumption. He analyzed the relative effect of strategies that use agents with different kill efficiencies by using a continuous-time stochastic birth–death multitype branching process model (Day 1986b) and simulation. The strategies he analyzed included alternating agents, interweaving but not strictly alternating strategies, and one-agent strategies. With each strategy a wide range of parameters were considered, including treatment-scheduling times, cell-doubling times, cell mutation rates, drug kill efficiencies, and single or cross resistance. The simulation results suggest two surprising conclusions: (1)When using two drugs it is best to use the least-effective drug first, or for a longer duration, and (2) for some values of tumor kinetics and drug kill parameters, nonalternating treatment can outperform alternation and combination treatment schedules. Practical application of these analyses depends upon knowing the appropriate drug kill parameters for each tumor of each patient, although simulation results provide guidelines in the absence of knowledge of exact parameter values.

4.3 Other Works and Applications

4.3.1 Fluctuation Analysis

An important application of a branching process involving mutations, similar to the model of Coldman and Goldie (Sect. 4.2), is the fluctuation test introduced by Luria and Delbrück in 1943 (Luria and Delbrück 1943). The model and some refinements will be considered in detail in Sect. 6.1. Here, we will describe the principle and provide some bibliography.

The progeny of a cell may exhibit a new trait that differs from their parent, and may pass on the new trait to their own progeny. Let us suppose that the change is due to a single irreversible mutation event. The mutation rate is expressed as the average number of mutations per cell division. Experimentally, a small number of cells are used to seed a series of independent cultures, cells in each culture are allowed to grow, and then the total number of cells in each culture is determined and also the number of mutant cells in each culture is determined. The number of cell divisions is estimated from the number of cells in each culture at the beginning and the end of the experiment.

Given parameter values, these models predict the distribution of the number of nonmutant and mutant cells at time t in a population started at time 0 by a single nonmutant cell. In particular, the following observable variables are of interest:

- $N(t)$, the expected total number of nonmutant and mutant cells at time t
- $r(t)$, the expected number of mutant cells at time t
- $P_0(t)$, the probability of mutant cells being absent from the population at time t

Conversely, given experimental values of $N(t)$, $r(t)$, and $P_0(t)$, it is possible to estimate the parameters of the models, in particular, mutation rates and probabilities.

Models in Sect. 6.1 illustrate how the estimates obtained differ if alternative assumptions are employed in addition to those originally used by Luria and Delbrück (1943). The literature of the subject includes many more refinements. A review of probability distributions of the number of mutants under differing assumptions can be found in Stewart et al. (1990). Ma et al. (1992) expanded these distributions into series involving discrete convolution powers. Cell death and differential growth rates were considered in a series of papers by Jones and co-workers (Jones et al.1994; Jones 1994).

Examples of applications, beyond the original data considered in Luria and Delbrück (1943), will be provided in Sect. 6.1. They mainly concern mutations to drug resistance in bacteria and cancer cells. One application in a different context is that by Hästbacka et al. (1992), who used a branching process of the Luria and Delbrück type to model the evolution of genetic disease and estimate the location of the disease gene.

An excellent review of various mathematical properties and approximations for the Luria and Delbrück distributions arising from the fluctuation analysis is provided in Angerer (2001).

4.4 Problems

1. *Cells with exponentially distributed lifetimes.* Consider the Markov time-continuous branching process with mean particle lifetime $1/\lambda$. Assume that at its death, each particle produces two specimens of progeny and that each of them survives, independently, with probability p. Find $h(s)$ and $F(s;t)$. *Consider the critical case separately.*
2. *Continued.* In the critical case, find the limit distribution of $\left\{\frac{Z(t;\omega)}{t}\middle| Z(t;\omega) > 0\right\}$, as $t \to \infty$. Compare the result with the corresponding general result for the Galton–Watson process. *Hint.* Consider the Laplace transform

$$\frac{F(e^{-\frac{u}{t}};t) - F(0;t)}{1 - F(0;t)}$$

   and use the results of the preceding problem.
3. *Explosions.* Consider the following branching process:
   - A single particle is born at $t = 0$. It lives 1 unit of time.
   - Each successive generation of particles lives three times shorter than the preceding one.
   - The pgf of progeny number in each generation is $f(s)$ such that $f'(1-) < \infty$.
   - Usual independence hypotheses are verified.

Find the pgf $F(s,t)$ of $Z(t)$, the number of particles present in the process at time $t \geq 0$. At what time the process may explode? What is the distribution of $Z(t)$ at that time? *Hint.* Consider separately the cases $f'(1-) \leq 1$ and $f'(1-) > 1$.

4. $\{X_n;\ n = 1, 2, \dots\}$ is a sequence of iid nonnegative random variables. Using the weak law of large numbers, demonstrate that

$$\lim_{n\to\infty} \mathrm{P}\{X_1 + X_2 + \cdots + X_n > t\} = 1, \quad \text{for any } t > 0.$$

   Hint. Assume first that $\mathrm{E}(X_1) < \infty$. If $\mathrm{E}(X_1) = \infty$, consider truncated rv's $Y_n = \min\{a, X_n\}$, where a is a constant.

5. *Clonal resistance revisited.* Consider the following version of the clonal resistance theory:
   a) In the process, there exist two types of particles, labeled 0 (sensitive) and 1 (resistant).
   b) The process is initiated by a single type 0 particle.
   c) The lifespans of particles are exponentially distributed independent random variables, with parameter λ.
   d) Each particle, at death, gives birth to exactly two progeny particles:
      – 0-particle produces either two 0-particles, wp $1 - \alpha$, or two 1-particles, wp α.
      – 1-particle produces two 1-particles.

Find the equations for the pgf's $F_0(s_0, s_1; t)$ and $F_1(s_1; t)$, see the lecture notes. Find and solve the equations for the expected counts of sensitive and resistant cells at time

t in the population started at time 0 by a single sensitive cell. Find and solve the equation for $P(t)$, the probability of no resistant cells at time t. Does the change in hypotheses alter the predictions of the theory?

6. *Serial mutations.* Consider the following branching process:
   a) In the process, there exist three types of particles, labeled 0, 1, and 2.
   b) The process is initiated by a single type 0 particle.
   c) The lifespans of particles are exponentially distributed independent random variables, with parameter λ.
   d) Each particle, at death, gives birth to exactly two progeny particles:
      – Each i-particle, $i = 0, 1$, produces either two i-particles, wp $1 - \alpha$, or one i particle and one $i + 1$-particle, wp α.
      – 2-particle produces two 2-particles.

The equations for the pgf's $F_0(s_0, s_1, s_2; t)$, $F_1(s_1, s_2; t)$, and $F_2(s_2; t)$ (the joint pgf's of the numbers of the 0-, 1-, and 2-particles, in the process initiated by a single 0-, 1-, and 2-particle, respectively; see the lecture notes) have the following form:

$$\dot{F}_0 = -\lambda F_0 + \lambda[\alpha F_0 F_1 + (1 - \alpha)F_0^2],$$

$$\dot{F}_1 = -\lambda F_1 + \lambda[\alpha F_1 F_2 + (1 - \alpha)F_1^2],$$

$$\dot{F}_2 = -\lambda F_2 + \lambda F_2^2,$$

with initial conditions $F_0(s_0, s_1, s_2; 0) = s_0$, $F_1(s_1, s_2; 0) = s_1$, and $F_2(s_2; 0) = s_2$. Find and solve the systems of equations for $P_1(t)$ and $P_2(t)$, the probabilities of no 1- and 2-cells at time t, respectively, in the process initiated by a single 0-particle. Draft the plots of $P_1(t)$ and $P_2(t)$. Conclusions?

7. Consider the time-continuous branching process with particle lifetimes distributed exponentially with expectation $1/\lambda$; started by a single ancestor. Assume that at its death, each particle produces two specimens of progeny with probability p and no progeny with probability $1 - p$.
   a) Find $h(s)$.
   b) In the critical case, find $F(s; t)$ and $P(t) = \mathrm{P}\{Z(t, \omega) = 0\}$.
   c) In the critical case, find the limit distribution of

$$\left\{ \frac{Z(t; \omega)}{t} \Big| Z(t; \omega) > 0 \right\},$$

   as $t \to \infty$. Compare the result with the corresponding general result for the Galton–Watson process. *Hint.* Consider the Laplace transform

$$\frac{F(\mathrm{e}^{-\frac{u}{t}}; t) - F(0; t)}{1 - F(0; t)}.$$

Chapter 5
The Bellman–Harris Process

The Bellman–Harris branching process is more general than the processes considered in the preceding chapters. Lifetimes of particles are nonnegative random variables with arbitrary distributions. It is described as follows. A single ancestor particle is born at $t = 0$. It lives for time τ which is a random variable with cumulative distribution function $G(\tau)$. At the moment of death, the particle produces a random number of progeny according to a probability distribution with pgf $f(s)$. Each of the first-generation progeny behaves, independently of each other and the ancestor, as the ancestor particle did, i.e., it lives for a random time distributed according to $G(\tau)$ and produces a random number of progeny according to $f(s)$. If we denote $Z(t)$ the particle count at time t,we obtain a stochastic process $\{Z(t), t \geq 0\}$. This so-called age-dependent process is generally non-Markov, but two of its special cases are Markov: the Galton–Watson process and the age-dependent branching process with exponential lifetimes. The Bellman–Harris process is more difficult to analyze, but it has many properties similar to these two processes.

Frequently, it is assumed that the distribution of lifetimes does not have an atom at $\tau = 0$, i.e., that $G(0+) = 0$, which is satisfied, among others, when the distribution has a density $g(\tau)$. This assumption prevents the process from producing infinitely many generations of particles in zero time. The assumption is not always required.

5.1 Integral Equations for the pgf and Basic Properties

5.1.1 Heuristic Derivations

We provide a heuristic derivation of the integral equation for the pgf of the number of particles in the Bellman–Harris process $Z(t)$. Since this is one of the most important equations in the theory of branching processes and since it has ramifications in some other branches of applied mathematics (renewal theory, deterministic population dynamics, and other), we also provide a complete derivation in the Appendix B.2.

Let us assume that the ancestor's lifetime is equal to τ. Then, for times $t < \tau$, the process consists of a single particle, the ancestor. For times $t \geq \tau$, the number

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_5

of particles in the process is equal to the sum of the numbers of particles in all subprocesses started by the first-generation progeny of the ancestor, i.e.,

$$Z(t) = \begin{cases} 1, & t < \tau, \\ \sum_{i=1}^{X} Z^{(i)}(t-\tau), & t \geq \tau, \end{cases}$$

where X is the number of the first-generation progeny of the ancestor, and $Z^{(i)}(t-\tau)$ are the iid copies of the process, started by these progeny particles at time τ. Denoting $F(t,s)$ the pgf of $Z(t)$, we obtain in terms of pgf's, conditional on τ

$$F(s,t) = \begin{cases} s, & t < \tau, \\ f[F(s,t-\tau)], & t \geq \tau. \end{cases} \tag{5.1}$$

Removing conditioning, i.e., integrating with respect to the distribution G, we obtain

$$F(s,t) = s[1-G(t)] + \int_{[0,t]} f[F(s,t-u)]dG(u). \tag{5.2}$$

This latter equation is identical to Eq. (B.6) in Appendix B.2.

In general, it is impossible to find explicit solutions of the integral Eq. (5.2). However, some special cases of interest are described by simpler equations.

Example 1 Galton–Watson process. *Suppose that $G(t) = \chi(t-T)$, where $\chi(t)$ is the unit step function at* 0*, i.e., that lifelengths of all particles are identical and equal to T.* Equation (5.1) *(as well as the integral* Eq. B.6*) assumes now the form*

$$F(s,t) = \begin{cases} s, & t \in [0,T), \\ f[F(s,t-T)], & t \in [T,\infty). \end{cases} \tag{5.3}$$

This implies that $F(s,t) = f_n(s)$ if $t \in [nT,(n+1)T)$ and also that $\{Z(nT),\ n = 0,1,\dots\}$ is a Galton–Watson process with progeny generating function $f(s)$.

Example 2 Markov age-dependent branching process. *If we consider the process with lifelength distributions $G(u) = 1-\exp(-\lambda u)$, i.e., the Markov age-dependent process, then the resulting integral equation can be differentiated side-by-side with respect to t, yielding (after some algebra) the differential* Eq. (4.6).

5.2 Renewal Theory and Asymptotics of the Moments

The theory of the renewal equation plays a major role in investigation of the asymptotic behavior of the Bellman–Harris process. The reason is that the moments of the process are solutions of renewal-type linear integral equations. However, the theory is also important for the *nonlinear* Eq. (5.2). We will follow Athreya and Ney (2004). Another source is the book by Feller (1971).

5.2.1 Basics of the Renewal Theory

Let us define the renewal function

$$U_m(t) = \sum_{n=0}^{\infty} m^n G^{*n}(t), \quad t \geq 0, \tag{5.4}$$

where G is a distribution on $[0, \infty)$, i.e., $G(t)$ is nonnegative, nondecreasing and $G(\infty) = 1$, and m is a positive constant. $G^{*n}(t)$ denotes n-fold convolution of function $G(t)$ by itself i.e., $G^{*n}(t) = \underbrace{G(t) * \cdots * G(t)}_{n\text{factors}}$, where $F(t) * G(t) = \int_{[0,t]} F(t-\tau) dG(\tau)$ and $F(t)$ and $G(t)$ are bounded nondecreasing functions on $[0, \infty)$.

Lemma 5.1 Athreya and Ney (2004). *If $mG(0+) < 1$, then $U_m(t) < \infty$ for all $t < \infty$, and is bounded on finite t-intervals.*

Let us consider the renewal equation

$$H(t) = \xi(t) + m \int_0^t H(t-y) dG(y), t \geq 0, \tag{5.5}$$

which also can be written as

$$H(t) = \xi(t) + m(H * G)(t),$$

where $\xi(t)$ is a given bounded measurable function on $[0, \infty)$ and $H(t)$ is the unknown function. Let $(\xi * U_m)(t) \equiv \int_0^t \xi(t-y) \mathrm{d}U_m(y)$ be the convolution of ξ and U_m which is well defined since U_m is nondecreasing and bounded.

Lemma 5.2 *$H \equiv \xi * U_m$ is the unique solution of* Eq. (5.5) *which is bounded on finite intervals.*

The following theorem can be found in Feller's (1971) book: We call a distribution a lattice, if its points of increase (or atoms of the corresponding probability measure) occupy isolated points separated by distances being integer multiples of a positive number a. Let us notice that if a distribution is nonlattice, then $G(0+) < 1$ is satisfied. The definition of direct Riemann integrability can be found in Feller (1971).

Theorem 5.1 *Assume $m = 1$ and let $\gamma = \int_0^\infty t \mathrm{d}G(t) \leq \infty$.*

1. *If $\xi_0 = \lim_{t\to\infty} \xi(t)$ exists, then the solution of* Eq. (5.5) *satisfies*

$$\lim_{t\to\infty} \frac{H(t)}{t} = \frac{\xi_0}{\gamma}. \tag{5.6}$$

2. *If ξ is directly Riemann integrable and $G(t)$ is nonlattice, then*

$$\lim_{t\to\infty} H(t) = \frac{1}{\gamma} \int_0^\infty \xi(y) \mathrm{d}y. \tag{5.7}$$

Definition 5.1 *The* Malthusian parameter *for m and G is the root,1,0,1unique provided it exists, of the equation*

$$m \int_0^\infty \mathrm{e}^{-\alpha y} \mathrm{d}G(y) = 1. \tag{5.8}$$

We denote it by $\alpha = \alpha(m, G)$.

Let us note that when $m \geq 1$, the Malthusian parameter always exists and is nonnegative. If $m < 1$, then α may not exist (if it does, it is negative).

When the Malthusian parameter exists, we can multiply Eq. (5.5) by $\mathrm{e}^{-\alpha t}$, and letting

$$H_\alpha(t) = \mathrm{e}^{-\alpha t} H(t); \;\; \mathrm{d}G_\alpha(t) = m\mathrm{e}^{-\alpha t} \mathrm{d}G(t); \;\; \xi_\alpha(t) = \mathrm{e}^{-\alpha t} \xi(t),$$

we obtain

$$H_\alpha(t) = \xi_\alpha(t) + \int_0^t H_\alpha(t - y) \mathrm{d}G_\alpha(y), t \geq 0. \tag{5.9}$$

Based on the above, part 2 of Theorem 5.1 can be used to obtain results of the following type:

Theorem 5.2 *If the Malthusian parameter α(m,G) exists, if* $\mathrm{e}^{-\alpha t} \xi(t)$ *is directly Riemann integrable, and if G is nonlattice and* $mG(0+) < 1$, *then the solution H of* Eq. (5.5) *satisfies*

$$H(t) \sim \mathrm{e}^{\alpha t} \left\{ \int_0^\infty \mathrm{e}^{-\alpha y} \xi(y) \mathrm{d}y \right\} \left\{ m \int_0^\infty y \mathrm{e}^{-\alpha y} \mathrm{d}G(y) \right\}^{-1}. \tag{5.10}$$

5.2.2 The Moments

In order to derive an equation for the expected number of particles in the process

$$\mathrm{E}[Z(t)] = \mu(t) = \frac{\partial F(s,t)}{\partial s}_{|s=1},$$

it is necessary to justify differentiation under the integral in Eq. (5.2). When this is accomplished, the following result is obtained:

Theorem 5.3 *Suppose* $mG(0+) < 1.E[Z(t)] \equiv \mu(t)$ *is the unique solution of*

$$\mu(t) = [1 - G(t)] + m \int_0^t \mu(t - y) \mathrm{d}G(y), \tag{5.11}$$

which is bounded on finite t-intervals.

Differentiating the pgf $F(s,t)$ more than once with respect to s, one obtains equations of similar type for higher moments of $Z(t)$. The equation for $[Z(t)] \equiv \mu(t)$

is of the renewal type. Theorem 5.2, applied to Eq. (5.11) yields the following asymptotic result:

Theorem 5.4 *Suppose* $mG(0+) < 1$.

1. *If* $m = 1$, *then* $\mu(t) = 1$.
2. *If* $m > 1$ *and* G *is nonlattice, then*

$$\mu(t) \sim ce^{\alpha t}, t \to \infty, \tag{5.12}$$

where α *is the Malthusian parameter for* (m, G) *and*

$$c = \frac{\int_0^\infty e^{-\alpha y}[1 - G(y)]dy}{m\int_0^\infty ye^{-\alpha y}dG(y)} = \frac{m-1}{\alpha m^2 \int_0^\infty ye^{-\alpha y}dG(y)}. \tag{5.13}$$

3. *If* $m < 1$, *if the Malthusian parameter* $\alpha(m, G)$ *exists, and if* $\int_0^\infty ye^{-\alpha y}dG(y) < \infty$, *then* (5.12) *and* (5.13) *hold, with* $\alpha < 0$.

5.3 Asymptotic Properties of the Process in the Supercritical Case

In the supercritical case, when the Malthusian parameter exists, the asymptotic behavior of the Bellman–Harris process is similar to the behavior of its expected value $\mu(t)$ and to the behavior of the Galton–Watson process. We define the random variable $W(t)$

$$W(t) = \frac{Z(t)}{n_1 e^{\alpha t}}, n_1 = \frac{m-1}{\alpha m^2 \int_0^\infty ue^{-\alpha u}dG(u)}, \tag{5.14}$$

where α is the Malthusian parameter (c.f. Definition 5.1). We see that $E[W(t)] \longrightarrow 1$, as $t \to \infty$ (c.f. Theorem 5.4).

Theorem 5.5 Athreya and Ney 2004. *Suppose that* $m > 1, f''(1-) < \infty, mG(0+) < 1$, *and* G *is not a lattice distribution. Then* $W(t)$ *converges with probability 1 and in mean squares to a random variable* W, *as* $t \to \infty$, *and*

$$E(W) = 1, \tag{5.15}$$

$$Var(W) = \frac{[m + f''(1-)]\int_0^\infty e^{-2\alpha u}dG(u) - 1}{1 - m\int_0^\infty e^{-2\alpha u}dG(u)}. \tag{5.16}$$

The variance of W *is positive.*

5.4 Application: Analysis of the Stathmokinetic Experiment

5.4.1 Age Distributions

It is frequently necessary to consider the number of particles (objects) not only in the whole process, $Z(t)$, but also in variously defined subsets of the process. Examples in the field of cell proliferation can be found in the review by Yanev (2010). The scope of that paper overlaps with the contents of the present section.

Suppose that for each object in the process, the lifetime τ is the sum of two independent random variables τ_1 and τ_2. This implies $G = G_1 * G_2$, where G_i is the distribution function of τ_i. More specifically, let us assume that the object's life is composed of *phase* 1 followed by *phase* 2, with respective durations τ_1 and τ_2. Suppose that we are interested in the number $X(u, t, \omega)$ of objects, at time t, which are in *phase* 1 and which have time $> u$ remaining to leave *phase* 1.

An analog of Eq. (5.2) is satisfied by the the pgf $F(s; u, t) = \mathrm{E}[s^{X(u,t)}]$

$$F(s; u, t) = \int_{0-}^{t+} f[F(s; u, t-\tau)]\mathrm{d}G(\tau) + s[1 - G_1(t+u)] + [G_1(t+u) - G(t)], \tag{5.17}$$

where $t, u \geq 0$ and $s \in [0, 1]$. Equation (5.17) is of the same type as Eq. (5.2). For a derivation, see Kimmel (1985).

5.4.2 The Stathmokinetic Experiment

The stathmokinetic experiment was employed by researchers to estimate parameters of cell cycle kinetics (see Darzynkiewicz et al. 1986, for a review). Basic notions concerning the cell cycle and cell cycle kinetics are explained in Sect. 2.2. When cell division is blocked by an agent that prevents completion of mitosis, the cells gradually accumulate in mitosis, emptying cells in the postmitotic phase G_1 as well as cells in the S phase (Fig. 5.1). The pattern of accumulation in mitosis (M) depends on the kinetic parameters of the cell cycle and is used for estimation of these parameters.

The following experimental law is observed in exponentially growing cell populations: Suppose that the cell population grows exponentially as $\mathrm{e}^{\lambda t}$. Let us define the *collection function* $g(t)$

$$g(t) = \ln[1 + f_M(t)], \tag{5.18}$$

where $f_M(t)$ is the fraction of the cells in mitosis at time t after starting the experiment. It is frequent that the initial portion of the graph of $g(t)$ is a straight line, the slope of which is equal to λ (Fig. 5.2). Based on this, the growth-rate parameter λ, inversely

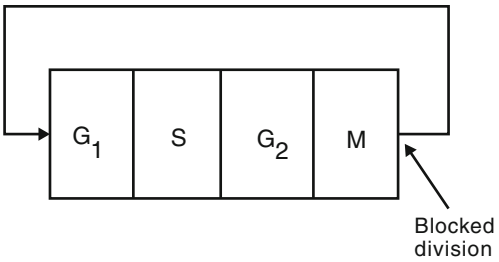


Fig. 5.1 Generally accepted subdivision of the cell cycle. After division, the daughter cells enter phase G_1, then traverse the phases S, G_2, and M, and then divide. The residence times in all the phases are treated as random. In the stathmokinetic experiment, the divisions are blocked, so that all the cells finally accumulate in M. (Source: Kimmel 1985)

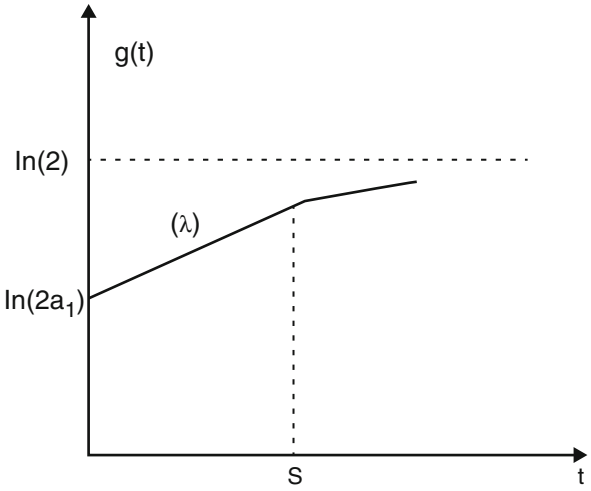


Fig. 5.2 Typical collection function $g(t)$. S is the minimum residence time in phase 1. For times from the interval $[0, S]$, the collection function is linear with slope λ. (Source: Kimmel 1985)

related to the duration of the cell cycle, can be found in an experiment of relatively short duration, in which only the fraction of cells in mitosis is followed. In more sophisticated versions of the stathmokinetic experiment, using the technique called flow cytometry, it is possible to follow fractions of cells in all cell cycle phases, and consequently to estimate more parameters of the cell cycle.

We present a model of the cell cycle based on the Bellman–Harris process. Based on this model, we derive a method of analysis of the stathmokinetic experiment which is independent of the particular functional form of the cell lifetime distribution. The approach follows Kimmel and Traganos (1986) and it is based on previous work by, among others, Jagers and Staudte.

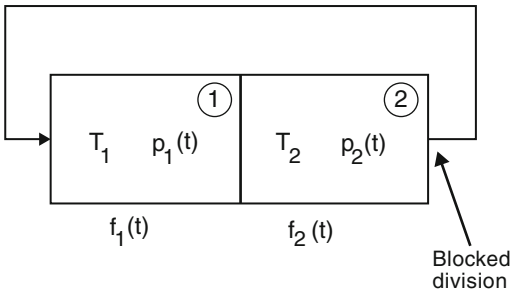


Fig. 5.3 Cell cycle subdivision into two "phases". T_i is the random residence time in phase i ($i = 1, 2$), p_i is its distribution density, $f_i(t)$ is the fraction of the initially cycling cells that are present in phase i at time t after the experiment is started. (Source: Kimmel 1985)

5.4.3 Model

It is assumed that proliferating cells follow the rules of a Bellman–Harris process with progeny pgf $f(s) = s^2$. The lifetimes of cells are iid rv's with a generic name T. T is assumed to be equal to the sum of two independent rv's T_1 and T_2, with densities p_1 and p_2, respectively (Fig. 5.3). After mitosis, each of the progeny cells enters *phase* 1, staying there for the random time T_1. Upon leaving phase 1, the cell enters *phase* 2 with duration T_2. Then the cell divides and both progeny reenter phase 1. It is assumed that the cells will have been growing for a long time in constant conditions before the beginning of the experiment (time $t = 0$), when cell divisions will be blocked by the application of a chemical agent.

After t_0 when the divisions have been blocked, the total number of cells stays unchanged but the transition from phase 1 to phase 2 continues. Therefore, the number of cells remaining in phase 1 at time t after t_0, is equal to $X(t, t_0)$, as defined in Sect. 5.4.1. The fraction $f_1(t)$ defined as

$$f_1(t) = \frac{\mathrm{E}[X(t, t_0)]}{\mathrm{E}[Z(t_0)]}, \tag{5.19}$$

where $Z(t_0)$ is the number of cells present at time t_0 (i.e., the number of objects in the Bellman–Harris process), is called the *exit curve* from phase 1. The *accumulation curve* in phase 2 is simply $f_2(t) = 1 - f_1(t)$.

Proposition 5.1 *The exit curve from phase 1 has the asymptotic form*

$$f_1(t) = \lim_{t_0 \to \infty} \frac{E[X(t, t_0)]}{E[Z(t_0)]} = 2[1 - P_1(t) - \alpha_1(t)], \tag{5.20}$$

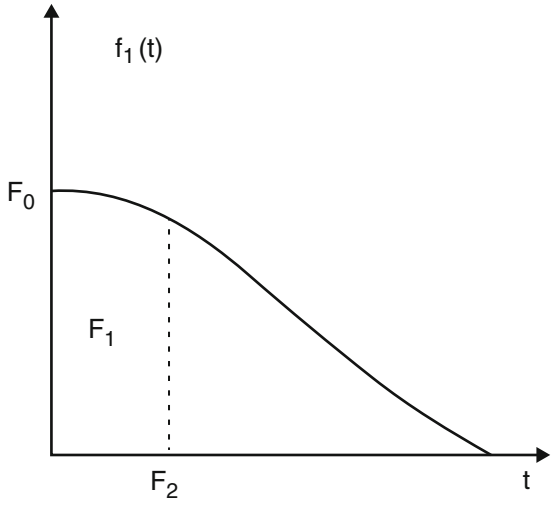


Fig. 5.4 Typical $f_1(t)$ exit curve. F_0 is the *exponential steady state (ESS)* cell fraction in phase 1; F_1 is the area under $f_1(t)$; F_2 is the coordinate of the mass center of the graph. (Source: Kimmel 1985)

where $P_1(t)$ is the tail distribution function $P[T_1 > t]$ of rv T_1 and

$$\alpha_1(t) = \mathrm{e}^{\lambda t} \int_t^\infty p_1(u)\mathrm{e}^{-\lambda u}\mathrm{d}u.$$

Here, λ is the Malthusian parameter being the unique real root of the equation

$$2\int_0^\infty \mathrm{e}^{-\lambda y}\mathrm{d}(P_1 * P_2)(y) = 1.$$

A detailed proof of Proposition 5.1 can be found in Kimmel (1985). Briefly, to find asymptotics of $f_1(t)$, we have to find the asymptotics of $\mathrm{E}[X(t, t_0)]$, as $t_0 \to \infty$. This is done by finding the integral equation for $\mathrm{E}[X(t, t_0)]$, which in turn is done by employing the pgf Eq. (5.17). Then, application of the renewal Theorem 5.2 yields the required asymptotics. An alternative proof could be carried out by using an appropriate random characteristic and Eq. (C.2).

The following two corollaries describe properties of the exit and collection functions (Fig. 5.4). Proofs can be found in Kimmel (1985): First corollary shows that first two moments of the random duration T_1 can be found as solutions of equations involving quantities F_0, F_1, and F_2, which can be computed from the graph of the exit function $f_1(t)$. Also, it shows how to invert the relationship (5.20) in order to compute the tail distribution of T_1, given the exit function. Second corollary demonstrates that the Malthusian parameter is equal to the slope of the linear portion of the accumulation curve $g(t)$.

Corollary 5.1 *Suppose the density p_1 exists and is bounded and that its two first moments exist. Then*

$$F_0 \equiv f_1(0) = 2(1-q), \tag{5.21}$$

$$F_1 \equiv \int_0^\infty f_1(u)\mathrm{d}u = 2\left[E(T_1) - \frac{1-q_1}{\lambda}\right], \tag{5.22}$$

$$F_2 \equiv \left[\int_0^\infty u f_1(u)\mathrm{d}u\right] \Big/ F_1 = \left[E(T_1^2) - \frac{2}{\lambda}E(T_1) + \frac{2}{\lambda^2}(1-q_1)\right] \Big/ F_1, \tag{5.23}$$

$$P_1(t) = 1 - \frac{f_1(t) - \frac{\mathrm{d}}{\mathrm{d}t} f_1(t)/\lambda}{2}, \tag{5.24}$$

where $q_1 = \alpha_1(0)$. The exit curve is assumed in its asymptotic form as in Proposition 5.1.

Corollary 5.2 *$g(t) = \lambda t + \ln(2q_1), t \le T_{\min}$, if and only if $p_1(t) = 0,\ t \le T_{\min}$.*

5.4.4 Estimation

Corollary 5.2 provides means of estimation of the Malthusian parameter of population growth. The parameter, λ, is equal to the slope of the initial straight line interval of the collection curve $g(t)$.

Knowing λ, and having the exit curve data for phase 1 (i.e., the values $f_1(0), f_1(t_1), \ldots, f_1(t_n)$, for a sequence of time points $0, t_1, \ldots, t_n$), it is possible to employ Corollary 5.1 to estimate the duration of phase 1. It can be done in two ways.

1. To calculate from data the "moments" F_0, F_1, and F_2 of the exit curve and solve the three first equations in Corollary 5.1 for $\mathrm{E}(T_1)$, $\mathrm{E}(T_1^2)$, and q_1.
2. To use the last equation in Corollary 5.1 to construct a nonparametric estimate of the cumulative distribution $P_1(t)$, based on experimentally recorded values $f_1(0), f_1(t_1), \ldots, f_1(t_n)$, and approximated values $\frac{d}{dt} f_1(0), \frac{d}{dt} f_1(t_1), \ldots, \frac{d}{dt} f_1(t_n)$.

A further discussion of applicability of these two methods can be found in Darzynkiewicz et al. (1986) and Kimmel (1985).

The decomposition of the cell cycle into abstract phases 1 and 2 can be carried out in various ways enabling analysis of stathmokinetic data in various biological compartments of the cell cycle (Fig. 5.1). Figure 5.5 depicts the estimation of the first moments of transit times through phases of the cell cycle of the Friend erythroleukemia cells (Kimmel 1985). Figure 6.10a (see Chap. 6) depicts the estimation of the distributions of transit times.

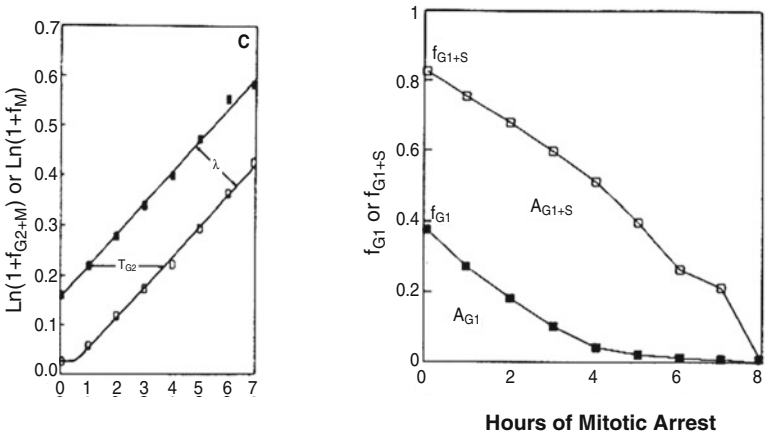


Fig. 5.5 Analysis of stathmokinetic data for cultured Friend erythroleukemia cells. First panel depicts the collection curves in phases M and $G_2 + M$. The (identical) slopes of the straight line portions of these curves provide the estimate of the Malthusian parameter $\lambda = 0.062$. Second panel depicts the nonparametric estimation of the mean duration of G_1 and S phase. G_1 (closed squares) and $G_1 + S$ (open squares) exit data are depicted in linear scale. The estimate of the mean duration of G_1 is calculated as $\mathrm{E}(T_1) = (A_{G_1} + f_{G_1}/\lambda)/2 = (0.82 + 0.38/0.062)/2 = 3.5$ h, where $f_{G_1} = 0.38$ is the fraction of G_1 cells at the beginning of stathmokinesis, and $A_{G_1} = 0.824$ is the area under the G_1 exit curve computed from the graph above based on piecewise linear approximation of the data. The estimate of the mean duration of $G_1 + S$ is calculated as $\mathrm{E}(T_1 + T_S) = (A_{G_1+S} + f_{G_1+S}/\lambda)/2 = (3.83 + 0.83/0.062)/2 = 8.6$ h, where $f_{G_1} = 0.83$ is the fraction of $G_1 + S$ cells at the beginning of stathmokinesis, and $A_{G_1+S} = 3.83$ is the area under the $G_1 + S$ exit curve. Subtraction $\mathrm{E}(T_S) = \mathrm{E}(T_1 + T_S) - \mathrm{E}(T_1) = 8.6 - 3.5 = 5.1$ h provides the estimate of S transit time. (Source: Kimmel 1985)

5.5 Other Works and Applications

5.5.1 Cell Populations

Cell populations are among natural objects that can be modeled using branching processes and this explains the great number and variety of publications devoted to this subject. A uniform presentation is difficult since different authors employed branching processes at different levels of generality or even branching processes disguised as deterministic models. The following account is chronological:

One of the earliest papers reviewing stochastic approaches to cell kinetics is Jagers (1983). Essentially, this is a treatment using general branching processes counted by random characteristics (Sect. C.1.2). Using this approach it is possible to provide a condensed mathematical description as well as to use the asymptotics of the supercritical process to describe the exponential growth of a population. The review also includes models with periodically varying coefficients and one of the earliest rigorous treatments of the stathmokinetic experiment (Sect. 5.4).

Another theoretical approach, quasi-stochastic, i.e., limited to expected-value processes is the paper by Staudte et al. (1984). It concerns models of regulatory mechanisms of the cell cycle. As such, it may be considered a precursor of the approach treated in detail in Sect. 7.8.1. Papers by Cowan (1985) and Cowan and Morris (1986) belong to the tradition of modeling of the cell cycle using a Bellman–Harris process (also, see Jagers' 1975 book; Kimmel 1980a, b, 1983). To be strict, this should be a multitype process, since cells in different phases of the cell cycle should be considered separately. However, due to the cyclical nature of the problem, it is possible to consider the interdivision time as a convolution of the durations of the successive cell cycle phases. Technically, this is carried out in a way similar to that described in Sect. 5.4.3.

One of the practical problems for which a mathematical answer is required is how to relate the doubling time t_d of an exponentially growing population, i.e., the time interval needed to increase the mean number of cells by a factor of 2, to the expected lifelength $\mathrm{E}(T)$ of an individual cell. The exact relationship has the form

$$t_d = \frac{\ln 2}{\alpha},$$

where α is the Malthusian parameter defined as the positive root of the equation

$$m\,\hat{f}(\alpha) = 1,$$

and $\hat{f}(\alpha)$ is the Laplace transform of the density $f(t)$ of the cell lifelength (Cowan 1985). There is no direct functional relationship between t_d and $\mathrm{E}(T)$. Similarly, there is no direct functional relationship between fractions of cells residing in distinctive phases of the cell cycle and the residence times of cells in these phases. The paper by Cowan (1985) provides approximations of the doubling time and of the proportions of cells in different phases in the terms of moments of the lifelength of cells and of the times of residence in cell cycle phases. This leads to a greater insight into the theory and some simple formulae which account for the variability.

In a subsequent paper, Cowan and Morris (1986) extend the analysis to the case of cells having different lifelength distributions in subsequent generations and becoming quiescent with some probability (possibly different in each generation). This allows modeling of transient effects in differentiating tissues and also of the embryonic phase of organism's growth.

The short book by Knolle (1988) presents the basic ideas of cell proliferation and some mathematical models of population growth. The main application is a cell cycle model with periodic coefficients used for modeling of cancer chemotherapy.

Axelrod et al.(1993) and Gusev and Axelrod (1995) use simulation of branching models to quantify the persistence of cell cycle times of *ras*-oncogene-transformed and non-transformed cells over many generations. The experimental system includes primary colonies of cells and secondary colonies grown from cells collected from primary colonies. Persistence of cell cycle times is determined by heritability of colony sizes (number of cells per colony). The problem of heritability was subsequently studied in more mathematically oriented papers, see Sect. 6.9.1.

Taïb (1995) studied the functional equation of the form $y'(x) = ay(\lambda x) + b(x)$ which arises in limiting cases of branching models of cell populations. The solution, important for applications, also has an intuitive interpretation as the probability density function of an infinite sum of independent but not identically distributed random variables.

5.5.2 Cell Proliferation

In this section, several papers will be discussed which are based on branching processes and related approaches and which tackle various aspects of cell proliferation. Nonclassical approaches, including nonstationary behavior or nonstandard models are stressed, rather than just the most recent papers.

One of the interesting nonstationary phenomena is the growth transient, before the population starts growing asynchronously and exponentially, referred to as the lag phase. Olofsson and Ma (2011) developed a branching process model of a bacterial population with an initial lag phase. Based on the Bellman–Harris process model, they established approximations in order to facilitate parameter estimation. Validity of the approximations and estimation procedures were tested using simulated data and found to be satisfactory.

A more complicated model has been developed by Nordon et al. (2011). This model is suitable for cells switching type, such as in differentiation of maturing progenitor cells in hemopoiesis. Cell proliferation and differentiation is described by a multitype branching Bellman–Harris process, a probability model that defines the inheritance of cell type. Cells first enter the G_1 phase and then proceed through the S, G_2, and M phases, and then divide. Cell type is defined by (1) a repression index related to the time required for S phase entry and (2) phenotype as determined by cell markers and division history. The inheritance of cell type is expressed as the expected number and type of progeny cells produced by a mother cell given her type. Expressions for the expected number and type of cells produced by a multicellular system are derived from the general model by making the simplifying assumption that cell generation times are independent. The specific form of the cell cycle model is the multitype Smith–Martin model (MSM; Smith and Martin 1973). It makes the assumption that cell generation times are distributed according to a shifted exponential distribution. The exponential part is the G_1 phase and the shift corresponds to the joint duration of the S, G_2, and M phases. Phenotype transitions are assumed to occur directly before entry into S phase. The authors convert the integral Bellman–Harris equations into a delay differential equation form, employing the Laplace transform. This has been accomplished in a more general form by Kimmel (1980a). The model of Nordon et al. (2011) has been tested using the data on the expansion of the number of human cord blood cells, positive for the surface protein CD34. The MSM model was fitted to cell division tracking data to identify cell cycle length, and the rates of CD34 antigen downregulation cell death (apoptosis). The authors also use the

inheritance-modified MSM (IMSM) model, which includes the influence of generation time memory so that mother and daughter generation times are correlated. They were able to fit the data with the model.

Orlando et al. (2009) applied the previously developed characterizing loss of cell cycle synchrony (CLOCCS) branching process model of cell population dynamics to time-series experiments in synchronized budding yeast *Saccharomyces cerevisiae* cells. The model accounts for cell cycle duration variability, asymmetric division, and distributed initial conditions. These features were also present in the previous model of Alexandersson (2001). A series of related quasi-stochastic models have been reviewed by Arino and Kimmel (1993). The Orlando et al. model provides a tool for "in silico synchronization" of the population and can be used to deconvolve population-level experimental measurements, such as temporal profiles of gene or protein expression. It also allows the direct comparison of assay measurements made in multiple experiments. The model can be fitted either to the yeast budding index or DNA content measurements, or both, and is adaptable to new forms of data.

The same group (Guo et al. 2012) also published an application of the model from Orlando et al. (2009) in a branching process deconvolution algorithm that provides a view of dynamic cell cycle processes, free from the convolution effects associated with imperfect cell synchronization. The method uses wavelet basis regularization, which increases the dynamic range of fit to data and the temporal resolution of time-series data. The method was applied to cell cycle time course data of transcription in the eukaryote *S. cerevisiae*. The algorithm made possible the identification of 82 genes transcribed almost entirely in the early G_1 part of the cell cycle.

It is interesting to note how mathematical descriptions of the same biological phenomenon evolve over time as new observations become available and emphasis shifts to even more detailed models. An example is provided by two stochastic approaches to modeling processes of proliferation in crypts of the small intestine and crypts of the colon. In an older paper, Loeffler and Grossman (1991) modeled intestinal epithelium with a two-level branching process. This is not unlike the branching-within-branching model of Kimmel (1997). The intestinal epithelium is one of the most rapidly regenerating tissues in mammals. Cell production takes place in the mouse intestinal crypts which contain about 250 cells. Only a minority of 1 out of 60 proliferating cells are able to maintain a crypt over a long period of time. The work is based on then available data about proliferation and extinction of cells in the crypts, and of the crypts themselves. The model assumptions are: (1) stem cells undergo a time-independent supercritical Markov branching process (Galton–Watson process), (2) a crypt divides if the number of stem cells exceeds a given threshold, and the stem cells are distributed to both daughter crypts according to binomial statistics, and (3) the size of the crypt is proportional to the stem cell number. This model described a new class of processes whose equilibrium and asymptotic behavior are contemplated. By comparison with crypt data available at the time they concluded that intestinal stem cells have a probability of over 0.8 of dividing asymmetrically producing one stem cell and one non-stem cell, and that the threshold number of stem cells is 8 or larger.

Tan and Yan (2010) developed a new stochastic and state space model for carcinogenesis of human colon cancer. They incorporated the biological mechanisms of chromosomal instability and microsatellite instability. The stochastic system was represented by two different pathways, one for each of the biological mechanisms. The observations, cancer incidence data, were represented by a statistical model. A generalized Bayesian approach was developed to estimate the parameters through the posterior modes of the parameters via Gibbs sampling procedures. The model was shown to fit the surveillance, epidemiology, and end results (SEER) age incidence data of human colon cancers from the National Cancer Institute of the National Institutes of Health, USA. Conclusions were drawn that could guide colon cancer prevention and control, and help predict future cancer cases. Comparison of these two models illustrates the profound change that has occurred in modeling in the past 20 years because of the access to more detailed biological information.

5.5.3 Estimation of Cell Lifetimes

Estimation of cell lifetimes can be carried out employing various consequences of the asymptotic balanced exponential growth of the supercritical Bellman–Harris process. The general principle is that the information accumulated in measurable characteristics of the cell population can be disentangled to extract the moments of cell lifetimes, probability of cell death, etc. One of the examples is the stathmokinetic experiment of Sect. 5.4, but other methods also can be used.

Jagers and Norrby (1974) proposed a method which involves sampling random cells from an exponentially growing population and following them to division or death. As the sampled cells will usually be of age greater than 0, the mean t_c of these times is less than the expected cell cycle duration T_c. Indeed

$$T_c = \frac{1-2p}{2(1-p)}(t_c - T_d/\ln 2),$$

where T_d is the doubling time of the population and p is the probability of cell death at division. Analogous expressions can be derived for variances. The authors provide statistics to estimate the moments of the cell cycle duration and provide examples of calculations for virally transformed and nontransformed human fetal cell lines. The conclusion is that the transformed cells have longer cell cycle times in spite of the population having a shorter doubling time.

The subject can be treated in more generality. If residence times in different cell cycle phases are random but not independent, then it is necessary to consider the following joint probabilities (Macdonald 1978):

$$\psi_i(u_1, \ldots, u_{i-1}, x, y, u_{i+1}, \ldots, u_p) \cdot \mathrm{d}u_1 \cdots \mathrm{d}u_{i-1} \cdot \mathrm{d}x \cdot \mathrm{d}y \cdot \mathrm{d}u_{i+1} \cdots \mathrm{d}u_p \quad (5.25)$$

that a cell chosen randomly from the population at time t is in phase i, has already spent times $u_1, \ldots, u_{i-1}$ in phases $1, \ldots, i-1$, time x in phase i, and is destined to

spend an additional time y in phase i and times $u_{i+1}, \ldots, u_p$ in phases $i+1, \ldots, p$. Although these probabilities are population dependent, the conditional distribution of $y + u_2 + \cdots + u_p$, given $i = 1$ and $x = 0$, is population independent and is the distribution of the lifelength of a newborn cell. On the contrary, both the backward lifelength $u_1 + \cdots + u_{i-1} + x$ and its forward counterpart $y + u_{i+1} + \cdots + u_p$, as well as their sum, are population dependent. Distributions $\psi_i(\cdot)$ can be computed under variety of assumptions. These include the exponential balanced growth, corresponding to the asymptotic behavior of a supercritical process, and also nonstationary cases (varying environment). A review is given by Macdonald (1978). Relationships of this kind allow construction of correct estimators of quantities more general than those considered by Jagers and Norrby (1974).

An analysis of estimation of mean cell cycle time, based on sample growth trajectories can be found in Hoel and Crump (1974).

Expression (5.25), in the balanced exponential growth version, was used by Cowan and Culpin (1981) to estimate the distribution function of cell residence times in subphases of the cell cycle. The experimental setup was a combination of in vivo fraction-labeled mitoses and arrested division (stathmokinesis) techniques. More specifically, chicken embryo cells were exposed to 5-bromodeoxyuridine (BUdR), which is incorporated into DNA during the S phase of the cell cycle. The amount of BUdR present in the cell is related to the number of times the cell underwent DNA synthesis (i.e., traversed the S phase) during the exposure. Just before the cells were removed from the embryo for measurement, colcemid or colchicine were injected to block further divisions. In this way, more cells accumulate in the prophase and metaphase (subphases of the M phase) in which the chromosomes can be resolved under the microscope. On the other hand, Macdonald's expression (5.25) makes it possible to calculate the expected numbers of cells that went through a given number of S phases and accumulated in prophase and metaphase. Using this expression, the model was fitted to observed cell counts, which allowed determining an optimum set of parameters characterizing the durations of cell cycle phases.

Different types of problems are considered in the papers by Axelrod and his coworkers. The main theme is estimation of parameters of cell cycle and the modes of dependence between related individuals, based on careful experiments with cell colonies, i.e., clonal cell populations. First of this series of papers (Axelrod et al. 1986) concerns the distributions of cell lifelengths of Friend experimental erythroleukemia cells. In addition to estimates of the α- and β-curves (tails of the distribution of lifelengths and of the distribution of differences between lifelengths of sib cells, respectively), the paper considers the issue of how dependencies between related cells are altered when cells are treated by cytotoxic agents (in this case, the differentiating agent dimethyl sulfoxide, DMSO). In Friend cells, the α-curves become more elongated (i.e., lifelenghts longer and more dispersed), while the β-curves are not altered. This is interpreted as consistent with sib–sib lifelengths correlations being increased in treated cells. In further papers, the main subject is the heterogeneity between colonies and its influence on estimated parameters such as correlations between lifetimes of related cells (Kuczek and Axelrod 1986). Kuczek

and Axelrod (1987) and Axelrod et al. (1997) introduce a divided colony assay to reduce the influence of heterogeneity on estimations of the influence of cytotoxic drugs on growth of cell colonies (also, see Axelrod and Kuczek 1989).

5.5.4 Bifurcating Autoregression

A particularly successful method of estimating parameters of cell proliferation is the bifurcating autoregression developed by Cowan and Staudte (1986). The method applies to branching populations with correlations between relatives defined in an autoregressive manner. In a genealogy of cells, if cell death is excluded, progeny of a cell m with generation time x_m can be labeled $2m$ and $2m+1$ and their generation times x_{2m} and x_{2m+1}, respectively. The ancestral cell is denoted 1. It is assumed that $x_1 \sim N(\mu, \sigma^2)$ and that given x_m the sib times x_{2m} and x_{2m+1} satisfy the relationships

$$\begin{aligned} x_{2m} - \mu &= \theta(x_m - \mu) + e_{2m}, \\ x_{2m+1} - \mu &= \theta(x_m - \mu) + e_{2m+1}, \end{aligned}$$

where (e_{2m}, e_{2m+1}) are bivariate normally distributed with common mean zero, common variance λ^2 and correlation coefficient ϕ. From these assumptions, the moments of x_m can be calculated, including the parent-progeny and sib–sib correlations. Consequently, a likelihood function of an observed pedigree can be calculated and numerically maximized to obtain the maximum likelihood estimates of μ, λ^2, and ϕ. The method was modified to accommodate relaxed assumptions and successfully employed to diverse data sets (Staudte 1992, Staudte et al. 1997 and references therein).

5.5.5 Branching Processes and Cancer Therapy

Some models of cancer chemotherapy aim to eradicate cancer stem cells while saving healthy stem cells. The paper by Sehl et al. (2011) uses an impressive array of analytical and computational tools to analyze a pair of stochastic processes describing proliferation and death of healthy stem cells and cancer stem cells exposed to chemotherapy. The question asked concerns the difference between these two types of cells in their birth and death rates that is required to eradicate the latter and preserve the former. Mutations, emergence of drug resistance, interactions of cancer and healthy cells and other complicating factors are disregarded. Because the biological model is simplified to the extreme, it allows effective mathematical analysis.

The mathematical approach used is quite sophisticated, for example, the distribution of the number of surviving healthy stem cells is considered at the random stopping time defined by extinction of the cancer stem cells. In addition, nontrivial applications of extreme value theory allow obtaining asymptotic distributions of the

times to extinction for cell populations started by multiple ancestors. Conclusions reached are interesting, even though it seems difficult to relate the simplified model to a relevant biological situation. It should be noted that if the framework of birth and death process is replaced by a more general Bellman–Harris branching process, the authors have to revert to simulations to reach any conclusions.

5.6 Problems

1. *Geometric Bellman–Harris process.* Suppose that the particle lifetime distribution is geometric, i.e., $\Pr\{\tau = i\} = (1-p)p^{i-1}$, $i \geq 1$ (progeny pgf is a general $h(s)$). Prove that $\{Z_i, i = 0, 1, \dots\}$, where $Z_i = Z(i)$, is a Galton–Watson process with some progeny pgf $f(s)$ (and consequently that $\{Z_i, i = 0, 1, \dots\}$ is Markov). Find $f(s)$. *Hint.* Write the equation for pgf of Z_i and proceed by induction. Another proof is possible using the lack of memory of the geometric distribution.
2. *The inverse problem.* Find the necessary and sufficient condition for a Galton–Watson process with progeny pgf $f(s)$ to be representable as a geometric Bellman–Harris process. *Hint.* Check if $h(s)$ corresponding to a given $f(s)$ is a pgf.
3. *Age distributions.* Find the integral equation for the pgf $F(s; y, t)$ of $Z(y, t)$ (number of particles at time t, with ages $\leq y$). Use the property

$$Z(y,t) = \sum_{k=1}^{X} Z^{(k)}(y, t-\tau), \text{if } \tau \leq t,$$

   and reasoning as in the heuristic derivation of Eq. (5.2).
4. *Expected age distributions.* Prove that if the Malthusian parameter exists, then, as $t \to \infty$, the normed expected age distribution $A(y,t) = \frac{\mathrm{E}[Z(y,t)]}{\mathrm{E}[Z(t)]}$ tends to the limit

$$A(y) = \frac{\int_0^y \mathrm{e}^{-\alpha z}[1 - G(z)]\mathrm{d}z}{\int_0^\infty \mathrm{e}^{-\alpha z}[1 - G(z)]\mathrm{d}z}.$$

   Hint. Find the integral equation for $\mu(y,t) \equiv \mathrm{E}\,[Z(y,t)]$ and use the asymptotics of Theorem 5.2.
5. *The $A \to B$ transition model of the cell cycle.* Suppose that in a proliferating cell population, a newborn cell, with probability p, stays dormant until it is prompted into further growth by a random "hit," which occurs (independently for each cell) with probability $\beta\tau + o(\tau)$ in any short time interval of duration τ. After this "hit," the cell requires fixed time T to grow and divide. Cells which do not require the "hit," start growing at the moment of birth. No cell death occurs. Find the limit age distribution $A(y)$. If, for a cell population growing long enough, the empirical age distribution can be found, can it help in establishing the value of p (which is a biologically important parameter)?

6. *Bellman–Harris process, the lattice case.* Consider the age-dependent branching process $\{Z_n, n = 0, 1, \dots\}$ with progeny pgf $h(s)$ and the lifetime distribution $\{g_i, i = 1, \dots, k\}$. Prove that the pgf $f_n(s)$ of Z_n is equal to $f_n^1(s, \dots, s)$, where $f_n^1(\mathbf{s})$ is the pgf of the k-type Galton–Watson process $\mathbf{Z}_n$ (initiated by a single type 1 particle), with the following progeny pgf's:

$$f^i(\mathbf{s}) = (1 - \gamma_i)h(s_1) + \gamma_i s_{i+1},\ i = 1, \dots, k-1,$$
$$f^k(\mathbf{s}) = (1 - \gamma_k)h(s_1),$$

where $\gamma_i = \frac{\tilde{G}_{i+1}}{\tilde{G}_i}$. (In other words, the total number of particles of all types in this k-type Galton–Watson process is equal to the number of particles in the lattice Bellman–Harris process. What is the interpretation of particle type here?)

7. *Perron–Frobenius root.* Assume $h'(1-) > 1$. Find the determinant equation for the maximum real eigenvalue ρ of matrix $\mathbf{M}$. Proceed by induction with respect to k. Check that the process is supercritical (i.e., that $\rho > 1$). Find the left eigenvector ν corresponding to ρ.
8. Show that the age distribution of particles in the process Z_n (i.e., the vector $(Z_{n1}, \dots, Z_{nk})$, where Z_{ni} is the number of particles with age i at time n), has pgf $f_n^1(\mathbf{s})$. Based on this and the limit law for the multitype supercritical positive regular Galton–Watson process 1,0,1, state the limit law for the age distribution of the lattice Bellman–Harris process.

Chapter 6
Multitype Processes

In the present chapter, we present models involving branching processes with many types of particles. Multitype models were sporadically employed in previous chapters. Here we offer a systematic treatment of asymptotic properties of the multitype Galton–Watson process in the supercritical case. However, we start with a motivating application, involving several multitype approaches to the fluctuation experiment analysis, which is one of the oldest but still useful tests of mutagenesis. Other applications follow.

6.1 Application: Mutations and Fluctuation Analysis

The progeny of a cell may exhibit a new trait that differs from their parent, and may pass on the new trait to their own progeny. Such a change is usually considered to be due to a single irreversible mutation event. However, a possibility exists that the observed change may be due to an event that has a finite probability of being reversible, or may be the result of more than one mutational event.

The rate at which mutations occur in populations of cells has been estimated using the fluctuation test introduced by Luria and Delbrück in 1943. The mutation rate is defined as the average number of mutations per cell division. Experimentally, a small number of cells is used to seed a series of independent cultures, the cells in each culture are allowed to grow in number, and then the total number of cells in each culture is determined and the number of mutant cells is determined in each culture. The number of cell divisions is estimated from the total number of cells in each culture at the beginning and at the end of the experiment. The mutation rate is then calculated in one of two ways (viz. from the total number of mutant cells or from the proportion of cultures with no mutant cells). An alternative method of calculating mutation rates is to use data observed at two successive generations (Niccum et al. 2012). Zheng (1999, 2005) has reviewed practical issues in estimating mutation rates resulting from the unique distribution of mutants per culture.

The two methods of calculating the mutation rate in the Luria and Delbrück fluctuation test do not always agree. This has motivated the investigation of models

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_6

that relax the assumptions of the original Luria and Delbrück fluctuation test—that changes in inherited traits are due to only one mutation event, and that each mutation is irreversible. Other possible deviations from the assumptions of the fluctuation test have been discussed (Foster 2006). One assumption made by Luria and Delbrück (1943) is that the total number of cells in each culture is the same. A robust estimator of mutation rates has been developed that takes into account the possibility of unequal numbers of total cells in different cultures (Wu et al. 2009). This is discussed in Sects. 6.1.8 and 6.1.9.

We present a series of models of cell growth and mutation. The purpose is to model the fluctuation experiment as applied to the analysis of data on drug resistance of cells. The material is based on the paper by Kimmel and Axelrod (1994). The classical fluctuation analysis is based on a model of cell proliferation and single-stage irreversible mutation introduced in Luria and Delbrück (1943). We summarize the hypotheses and predictions of that model and of four other models employing different hypotheses. These models are modifications of the Luria–Delbrück model, including random cell interdivision time, cell death, and two-stage mutations with the first stage being reversible.

Given the parameter values, these models predict the distribution of the number of nonmutant and mutant cells at time t, in a population started at time 0, by a single nonmutant cell. In particular, the following observable variables are of interest:

- $N(t)$, the expected total number of nonmutant and mutant cells at time t
- $r(t)$, the expected number of mutant cells at time t
- $P_0(t)$, the probability of mutant cells being absent from the population at time t

Conversely, given the experimental values of $N(t)$, $r(t)$, and $P_0(t)$, it is possible to estimate the parameters of the models, in particular, the mutation rates and probabilities.

6.1.1 Luria–Delbrück Model

The hypotheses are as follows (see Fig. 6.1a and Table 6.1):

1. Two types of cells exist in the population: *type 0* nonmutant cells and *type 1* mutant cells.
2. The population of cells has interdivision times equal to $\ln 2$.
3. Each cell, at the moment of division, gives birth to two daughter cells. The type of each of these daughters is the same as that of the mother cell.
4. During its lifetime, independently of any other events, a *type 0* cell undergoes an irreversible transformation into a *type 1* cell, with probability $a\Delta t + o(\Delta t)$, in any brief lifetime interval $(t, t + \Delta t)$. The constant a is called the *transition* or *mutation* rate. This implies that if the time from birth to mutation is denoted by T, then $\mathrm{P}[T > t] = \max[\exp(-at), \exp(-a \ln 2)]$, i.e., the mutation may not occur at all wp $\exp(-a \ln 2) = 2^{-a}$.

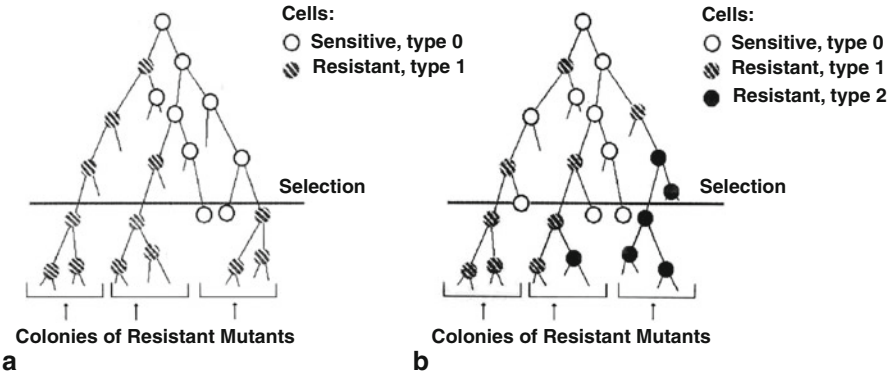


Fig. 6.1 Schematics of transitions admitted in **a** the one-stage models, and **b** the two-stage model. (Source: Kimmel and Axelrod 1994)

Table 6.1 Summary of hypotheses of the models considered

Model	Interdivision time	Probability of cell death	Number of stages	Probability of mutation
Luria–Delbrück	ln (2)	0	1	$a\Delta t$ in $(t, t+\Delta t)$ (irreversible)
Markov branching	1 (expected)	0	1	$a\Delta t$ in $(t, t+\Delta t)$ (irreversible)
Galton–Watson	1	0	1	α per daughter cell (irreversible)
Galton–Watson with cell death	1	δ	1	α per daughter cell (irreversible)
Galton–Watson two-stage	1	0	2	$0 \to 1: \alpha_{01}$ $1 \to 0: \alpha_{10}$ (reversible) $1 \to 2: \alpha_{12}$ (irreversible)

The analysis of the model, carried out originally in Luria and Delbrück (1943) and reworked in Lea and Coulson (1949), is based on the assumption that the population as a whole is large enough to be treated deterministically, while the mutation events are rare and therefore the mutants have to be counted in a probabilistic manner. Solutions, which were derived in Lea and Coulson (1949), are listed in the first row of Table 6.2. We do not provide derivations, but referring the reader to Kimmel and Axelrod (1994). The original estimate of the mutation rate in *Escherichia coli* bacterium obtained by Luria and Delbrück (1943), varied from $0.32 \times 10^{-8} - 2.45 \times 10^{-8}$ per bacterium per division. Further discussion of the mutation rates is found in Sect. 6.10.1.

Table 6.2 Expressions for the expected total count of cells $N(t)$, for the expected count of mutant cells $r(t)$, and for the probabilities of no mutant cells $P_0(t)$, in the models considered

Model	$N(t)$	$r(t)$	$P_0(t)$
Luria–Delbrück	e^t	ate^t	$\exp(-ae^t)$
Markov branching process	e^t	$e^t(1-e^{-at})$	$\frac{a+1}{ae^{(a+1)t}+1}$
Galton–Watson process	2^t	$2^t[1-(1-\alpha)^t]$	$(1-\alpha)^{2^{(t+1)}-2}$
Galton–Watson with cell death	$[2(1-\delta)]^t$	$[2(1-\delta)]^t[1-(\frac{1-\alpha-\delta}{1-\delta})^t]$	Eqs. (6.23)–(6.24)
Galton–Watson two-stage	2^t	$2^t - \frac{\rho_1^t}{A_1} - \frac{\rho_2^t}{A_2}$	Eqs. (6.25)–(6.27) and (6.33)

6.1.2 The Markov Branching Process Model

In this model, the interdivision time is not constant but is random with exponential distribution. Hypothesis 2 is therefore replaced by the following (c.f. Table 6.1 and Fig. 6.1a):

2. All cells in the population have *exponentially distributed* interdivision times with mean (expected) value equal to 1.

The distributions of the numbers of nonmutant and mutant cells are characterized by the following pgf's:

$$F_0(s_0, s_1; t) = \mathrm{E}[Z_0(t)^{s_0} Z_1(t)^{s_1} | Z_0(0) = 1, Z_1(0) = 0], \tag{6.1}$$

$$F_1(s_0, s_1; t) = \mathrm{E}[Z_0(t)^{s_0} Z_1(t)^{s_1} | Z_0(0) = 0, Z_1(0) = 1], \tag{6.2}$$

where $t \geq 0$, $s_1, s_2 \in [0, 1]$. $Z_0(t)$ (respectively, $Z_1(t)$) is the number of nonmutant (respectively, mutant) cells at time t. Function F_0 is the pgf of the population started by a single nonmutant cell, while function F_1 is the pgf of a clone started by a single mutant cell. We will write $F_i(s; t)$ or $F_i(t)$ instead of $F_i(s_0, s_1; t)$.

The model is a two-type age-dependent Markov branching process and the following differential equations are satisfied by the pgf's $F_0(t)$ and $F_1(t)$ (c.f., Sect. 4.2.1):

$$\frac{\mathrm{d}}{\mathrm{d}t} F_0(t) = -(a+1)F_0(t) + (a+1)\left[\frac{1}{a+1} F_0(t)^2 + \frac{a}{a+1} F_1(t)\right], \tag{6.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} F_1(t) = -F_1(t) + F_1^2(t), \quad t \geq 0. \tag{6.4}$$

The initial conditions are $F_i(t) = s_i$, $i = 0, 1$. The form of Eq. (6.3) and (6.4) can be understood by comparison with Eq. (4.10) and (4.11). Under the new Hypothesis 2, after a time, which is distributed exponentially with parameter $a + 1$, either two type 0 cells are produced [wp. $1/(a + 1)$] or a single type 1 cell [wp $a/(a + 1)$] is produced. This latter cell is a "type 1 continuation" of the type 0 cell.

We are interested in evaluating $N(t)$, $r(t)$, and $P_0(t)$. They can be expressed as,

$$N(t) = \mathrm{E}[Z_0(t) + Z_1(t) | Z_0(0) = 1, Z_1(0) = 0] = \left(\frac{\partial}{\partial s_0} + \frac{\partial}{\partial s_1} \right) F_0(s;t)_{|s_0=s_1=1}, \tag{6.5}$$

$$r(t) = \mathrm{E}[Z_1(t) | Z_0(0) = 1, Z_1(0) = 0] = \frac{\partial}{\partial s_1} F_0(s;t)_{|s_0=s_1=1}, \tag{6.6}$$

$$P_0(t) = F_0(1, 0; t). \tag{6.7}$$

Solving resulting equations yield results displayed in Table 6.2.

6.1.3 *The Galton–Watson Process Model*

In this model, cells mutate immediately following division. Hypotheses 3 and 4 are therefore replaced by the following (c.f. Table 6.1 and Fig. 6.1a):

3. Each cell, at the moment of division, gives birth to two daughter cells. The type of each of the daughters may or may not be the same as that of the mother cell.

4. Following division, a *type 0* daughter cell undergoes irreversible transformation into a *type 1* cell with probability α. The constant α is now called the transition or mutation *probability*.

The distributions of nonmutant and mutant cells can be characterized by the pgf's $F_0(s_0, s_1; t)$ and $F_1(s_0, s_1; t)$ as defined in Eqs. (6.1) and (6.2), except that the time variable t now assumes only nonnegative integer values, equal to the multiples of the interdivision time.

The pgf's $F_0(t)$ and $F_1(t)$ satisfy a system of recurrence equations, stemming from the following vector generalization of the backward iteration (3.2):

$$F(s,t) = h[F(s,t-1)],$$

where

$$F = (F_0, F_1), \quad h = (h_0, h_1),$$

$$h_0(s_0, s_1) = [(1-\alpha)s_0 + \alpha s_1]^2, \quad h_1(s_0, s_1) = s_1^2.$$

Substituting h_0 and h_1 as given above, we obtain

$$F_0(s;t) = [(1-\alpha)F_0(s;t-1) + \alpha F_1(s;t-1)]^2, \tag{6.8}$$

$$F_1(s;t) = [F_1(s;t-1)]^2, \tag{6.9}$$

where $s = (s_0, s_1)$, $t = 1, 2, \ldots$, with initial conditions $F_i(s;0) = s_i$, $i = 0, 1$. Recurrences (6.8) and (6.9) cannot be solved explicitly but differentiation side-by-side of (6.8) and (6.9) with respect to s_0 and setting $s_0 = s_1 = 1$ yields

$$\begin{aligned}\mathrm{E}[Z_0(t)|Z_i(0) = \delta_{0i}] =& 2\{(1-\alpha)\mathrm{E}[Z_0(t-1)|Z_i(0) = \delta_{0i}] \\ &+ \alpha\mathrm{E}[Z_0(t-1)|Z_i(0) = \delta_{1i}]\},\end{aligned} \tag{6.10}$$

$$\mathrm{E}[Z_0(t)|Z_i(0) = \delta_{1i}] = 2\mathrm{E}[Z_0(t-1)|Z_i(0) = \delta_{1i}],\ t = 1, 2, \ldots, \tag{6.11}$$

with initial conditions $\mathrm{E}[Z_0(0)|Z_i(0) = \delta_{0i}] = 1$, $\mathrm{E}[Z_0(0)|Z_i(0) = \delta_{1i}] = 0$. This yields $\mathrm{E}[Z_0(t)|Z_i(0) = \delta_{1i}] = 0$, $t = 0, 1, 2, \ldots$ and

$$\mathrm{E}[Z_0(t)|Z_i(0) = \delta_{0i}] = [2(1-\alpha)]^t,\ t = 0, 1, 2, \ldots, \tag{6.12}$$

as expected. Since there is no cell death assumed,

$$N(t) = \mathrm{E}[Z_0(t) + Z_1(t)|Z_i(0) = \delta_{0i}] = 2^t,\ t = 0, 1, 2, \ldots. \tag{6.13}$$

Equations (6.12) and (6.13) yield

$$r(t) = \mathrm{E}[Z_1(t)|Z_i(0) = \delta_{0i}] = 2^t - [2(1-\alpha)]^t,\ t = 0, 1, 2, \ldots, \tag{6.14}$$

as displayed in Table 6.2.

To obtain $P_0(t)$ we use the definition (6.7) and also denote $P_1(t) = F_1(1, 0; t)$. Substitution of $s_0 = 1$, $s_1 = 0$ in (6.8) and (6.9) yields,

$$P_0(t) = [(1-\alpha)P_0(t-1) + \alpha P_1(t-1)]^2, \tag{6.15}$$

$$P_1(t) = [P_1(t-1)]^2,\ t = 1, 2, \ldots, \tag{6.16}$$

with initial conditions $P_0(0) = 1$ and $P_1(0) = 0$. Therefore,

$$P_0(t) = (1-\alpha)^{2^{t+1}-2},\ t = 0, 1, 2, \ldots, \tag{6.17}$$

as displayed in Table 6.2.

6.1.4 *The Galton–Watson Process Model with Cell Death*

In this model, each of the daughter cells (mutant or nonmutant) may also die with some probability. Hypothesis 4 is therefore replaced by the following (c.f. Table 6.1):

4. Following division, a *type 0* daughter cell either undergoes irreversible transformation into a *type 1* cell with probability α, or dies with probability δ, or stays *type 0*

with probability $1-\alpha-\delta$. The *type 1* daughter cell may either die with probability δ or stay alive with probability $1-\delta$.

The presence of cell death leads to the following modification of Eqs. (6.8) and (6.9):

$$F_0(s;t) = [(1-\alpha-\delta)F_0(s;t-1) + \alpha F_1(s;t-1) + \delta]^2, \tag{6.18}$$

$$F_1(s;t) = [(1-\delta)F_1(s;t-1) + \delta]^2, \;\; t = 1,2,\ldots, \tag{6.19}$$

with initial conditions $F_i(s;0) = s_i$, $i = 0,1$. We obtain,

$$\mathrm{E}[Z_0(t)|Z_i(0) = \delta_{0i}] = [2(1-\alpha-\delta)]^t, \;\; t = 0,1,2,\ldots, \tag{6.20}$$

and

$$N(t) = \mathrm{E}[Z_0(t) + Z_1(t)|Z_i(0) = \delta_{0i}] = [2(1-\delta)]^t, \;\; t = 0,1,2,\ldots, \tag{6.21}$$

which yields

$$r(t) = \mathrm{E}[Z_1(t)|Z_i(0) = \delta_{0i}] = [2(1-\delta)]^t \left[1 - \left(\frac{1-\alpha-\delta}{1-\delta}\right)^t\right], \;\; t = 0,1,2,\ldots, \tag{6.22}$$

as displayed in Table 3.2. Substitution of $s_0 = 1$ and $s_1 = 0$ in (6.18) and (6.19) yields,

$$P_0(t) = [(1-\alpha-\delta)P_0(t-1) + \alpha P_1(t-1) + \delta]^2, \tag{6.23}$$

$$P_1(t) = [(1-\delta)P_1(t-1) + \delta]^2, \;\; t = 1,2,\ldots, \tag{6.24}$$

with initial conditions $P_0(0) = 1$, $P_1(0) = 0$, where $P_0(t)$ is the probability of no mutant cells at time t in the population derived from a nonmutant cell, while $P_1(t)$ is the extinction probability (by time t) of a clone started by a mutant. This recurrence has to be solved numerically.

6.1.5 Two-Stage Galton–Watson Process Model

In this model, two stages of mutant cells are present, *type 1* and *type 2*. Mutation from *type 0* to *type 1* is reversible, while mutation from *type 1* to *type 2* is irreversible. Hypothesis 4 is therefore replaced by the following (c.f. Table 6.1 and Fig. 6.1b):

4. Following division,
- A *type 0* daughter cell undergoes transformation into a *type 1* cell, with probability α_{01}.

- A *type 1* daughter cell undergoes a reverse transformation into a *type 0* cell, with probability α_{10}.
- A *type 1* daughter cell undergoes irreversible transformation into a *type 2* cell, with probability α_{12}.

The two-stage mutation model is a three-type Galton–Watson process. Its distributions are described by pgf's $F_i(s_0, s_1, s_2; t)$, $i = 0, 1, 2$, where F_0 is the joint pgf of the numbers of cells of types 0, 1, and 2 in the population started by a single nonmutant cell, F_1 is the joint pgf in the population started by a single stage-1 mutant cell, and F_2 is the pgf of the stage-2 mutant clone started by a single stage-2 irreversible mutant. The hypotheses of the model lead to the following recurrent equations for the pgf's:

$$F_0(s;t) = [(1-\alpha_{01})F_0(s;t-1) + \alpha_{01}F_1(s;t-1)]^2, \tag{6.25}$$

$$F_1(s;t) = [\alpha_{10}F_0(s;t-1) + (1-\alpha_{10}-\alpha_{12})F_1(s;t-1) + \alpha_{12}F_2(s;t-1)]^2, \tag{6.26}$$

$$F_2(s;t) = [F_2(s;t-1)]^2, \quad t = 1, 2, \ldots, \tag{6.27}$$

with initial conditions $F_i(s;0) = s_i$, $i = 0, 1, 2$. Let us denote $M(t) = (M_{ij}(t))_{i,j=0,1,2}$, the matrix of expected cell counts

$$M_{ij} = \mathrm{E}[Z_j(t)|Z_k(0) = \delta_{ik}, k = 0, 1, 2] = \frac{\partial F_i(1;t)}{\partial s_j}. \tag{6.28}$$

Differentiating system (6.25)–(6.27) we obtain

$$M(t) = \mu M(t-1), \quad t = 1, 2, \ldots, \tag{6.29}$$

where μ is the expected progeny matrix,

$$\mu = 2\begin{pmatrix} 1-\alpha_{01} & \alpha_{01} & 0 \\ \alpha_{10} & 1-\alpha_{10}-\alpha_{12} & \alpha_{12} \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.30}$$

The initial condition is $M(0) = I$ (the identity matrix). We obtain

$$M(t) = \mu^t, \quad t = 0, 1, 2, \ldots. \tag{6.31}$$

Involved but standard calculations consisting of finding the eigenvalues and eigenvectors of matrix μ lead to the following explicit expression for $r(t)$:

$$r(t) = M_{01}(t) + M_{02}(t) = 2^t - \frac{\rho_1^t}{A_1} - \frac{\rho_2^t}{A_2}, \quad t = 0, 1, 2, \ldots. \tag{6.32}$$

where

$$\rho_i = (2 - \alpha_{01} - \alpha_{10} - \alpha_{12}) + (-1)^i \sqrt{(\alpha_{10} + \alpha_{12} - \alpha_{01})^2 + 4\alpha_{01}\alpha_{10}},$$

and

$$A_i = 1 + \frac{[2(1 - \alpha_{01}) - \rho_i]^2}{4\alpha_{01}\alpha_{10}}, \quad i = 1, 2.$$

Recurrent equations for

$$P_0(t) = F_0(1, 0, 0; t), \tag{6.33}$$

are obtained from system (6.25)–(6.27) using substitution $s_0 = 1$, $s_1 = s_2 = 0$.

6.1.6 *The Single-Stage Models Versus Data*

The question considered in this section is whether the single-stage models can simultaneously reproduce the r and P_0 values obtained from experimental data. Each single-stage model yields, for a given value of mutation rate a or mutation probability α, and for a given sample size $N(t)$, a uniquely determined pair of values $r(t)$ and $P_0(t)$. We will call the $r - P_0$ plot the set of all such points in the $r - P_0$ plane. The equation of the $r - P_0$ plot can be found by eliminating a (or α) from the expressions for $r(t)$ and $P_0(t)$ in Table 6.2. For example, the $r - P_0$ plot of the Luria–Delbrück model has the following equation:

$$P_0 = \exp[-r/\ln(N)].$$

By graphing the experimentally obtained estimates of r and P_0 together with the corresponding $r - P_0$ plot for an appropriate N, we can verify whether the model can fit the data.

The first series of comparisons includes the original data on bacterial resistance to phage from Luria and Delbrück (1943), almost perfectly matched by the single-stage models. We present in Fig. 6.2, the $r - P_0$ plot and the data point of experiment 23 of Luria and Delbrück (1943). Notice that the data point interpolates between the models with constant lifetime and exponentially distributed lifetime, two extreme alternatives of lifetime distributions. Experiment 22 provides a similar match. For these classical data, the single-stage models seem perfectly satisfactory.

As a contrast, we analyze the gene amplification data from Tlsty et al. (1989) and Murnane and Yezzi (1988; details in Kimmel and Axelrod 1994).

Figures 6.3 and 6.4 demonstrate that the r and P_0 values obtained in this way are not matched by the $r - P_0$ plots of the Luria–Delbrück, Galton–Watson, and Markov branching process models (from left to right). The overall tendency of these three models is to overestimate either r or P_0. Taking into account cell death, makes the match even worse (Fig. 6.5).

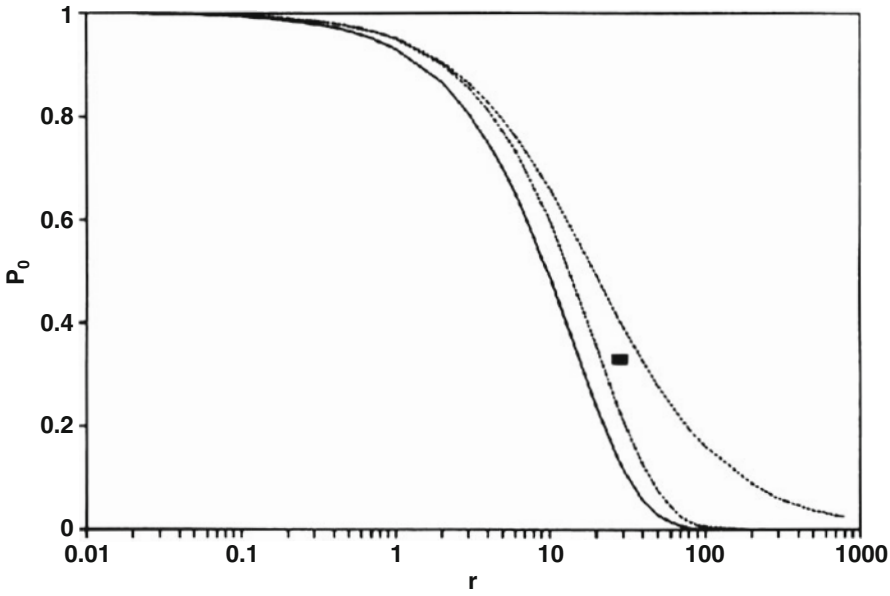


Fig. 6.2 Bacterial phage resistance data from Experiment 23 of (Luria and Delbrück 1943), and the $r - P_0$ plots ($N = 2.4 \times 10^8$) of the Galton–Watson, Luria–Delbrück and Markov branching process models. (Source: Kimmel and Axelrod 1994)

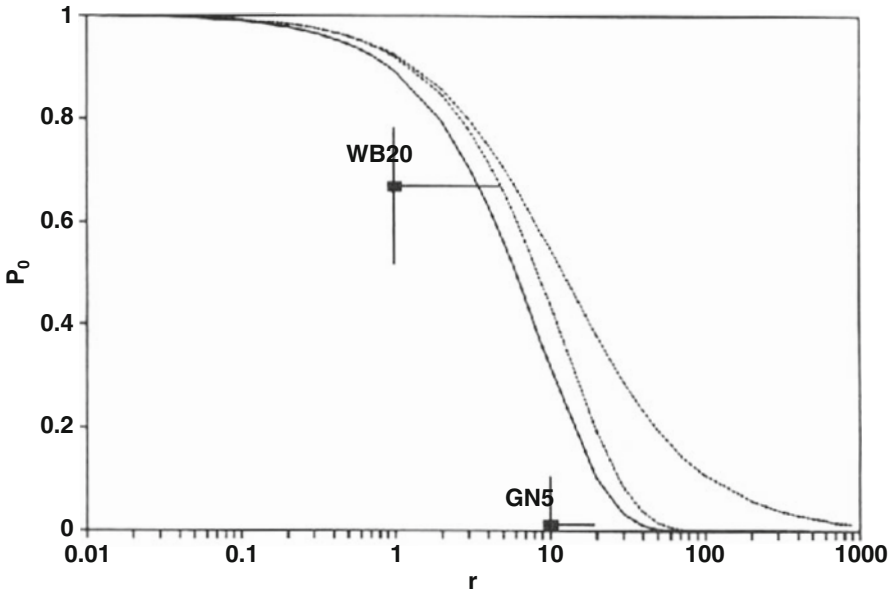


Fig. 6.3 The gene amplification data for WB_{20} and GN_5 cells from (Tlsty et al. 1989), and the $r - P_0$ plots ($N = 2 \times 10^5$) of the Galton–Watson, Luria–Delbrück, and Markov branching process models (from left to right). (Source: Kimmel and Axelrod 1994)

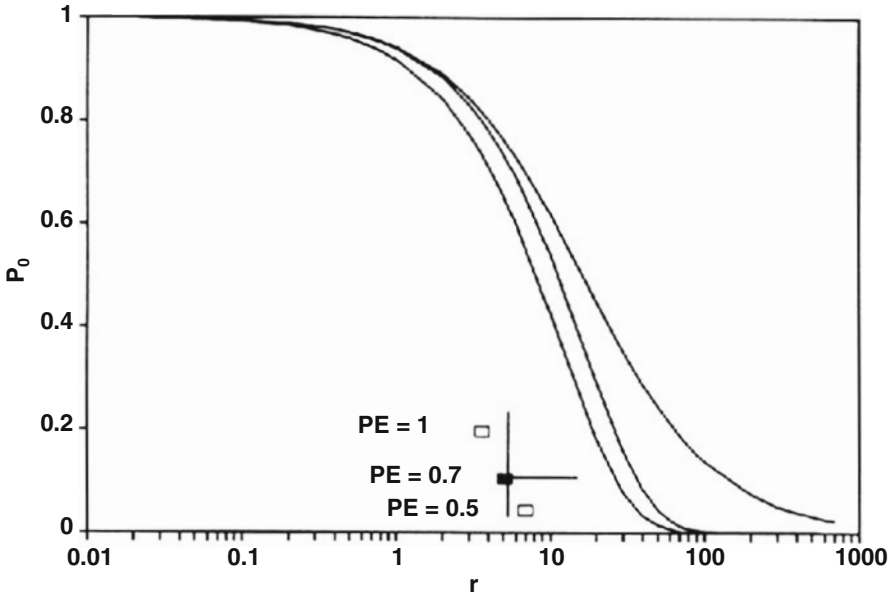


Fig. 6.4 The gene amplification data for LM205 cells from (Murnane and Yezzi 1988), and the $r-P_0$ plots ($N = 1.1 \times 10^7$) of the Galton–Watson, Luria–Delbrück and Markov branching process models. The three experimental points were obtained using three different hypothetical values of plating efficiency. (Source: Kimmel and Axelrod 1994)

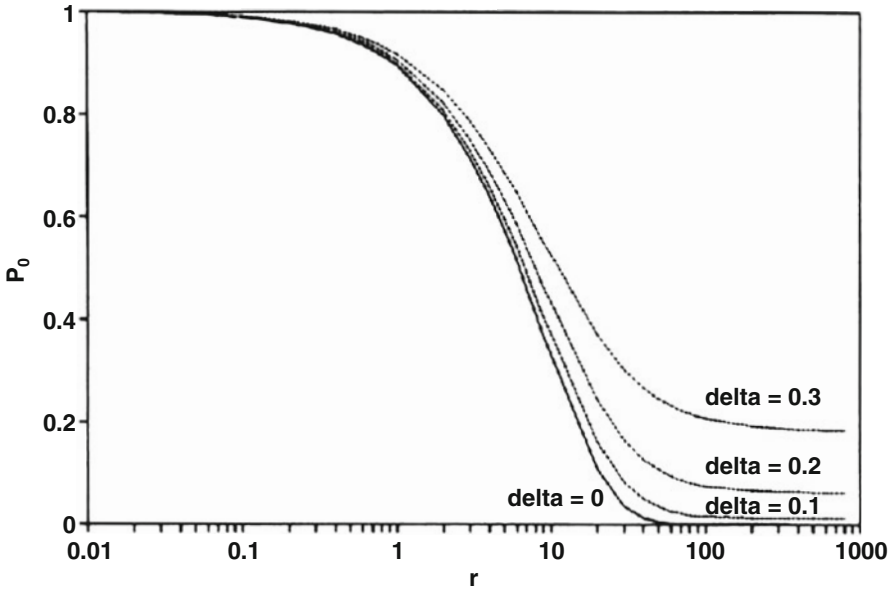


Fig. 6.5 The effect of cell death on the $r - P_0$ plots of the Galton–Watson model with cell death ($N = 2 \times 10^5$). Delta is the probability of cell death. (Source: Kimmel and Axelrod 1994)

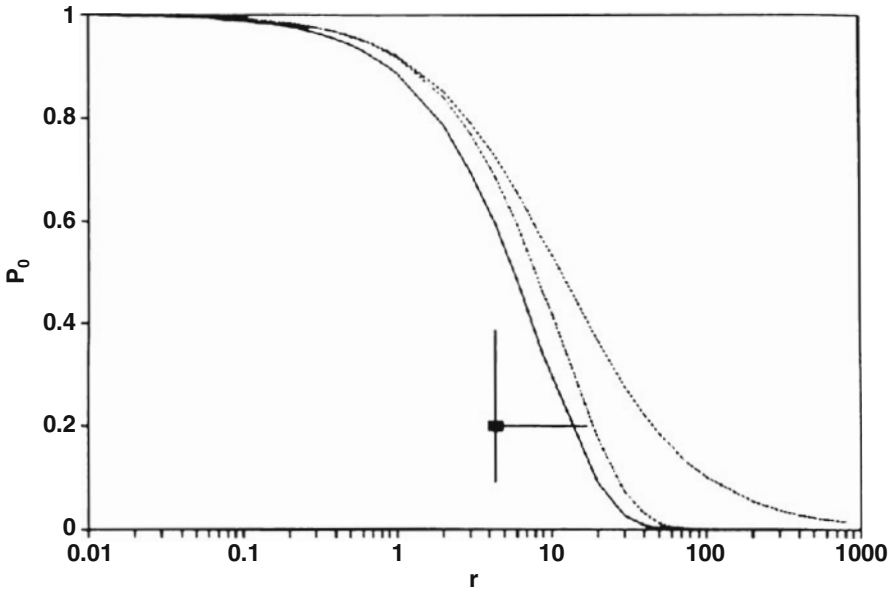


Fig. 6.6 The drug resistance data for $M - Mc$ mouse cells from experiment 1 of (Morrow 1970), and the $r - P_0$ plots ($N = 10^5$) of the Galton–Watson, Luria–Delbrück, and Markov branching process models (from left to right). (Source: Kimmel and Axelrod 1994)

Figures 6.6 and 6.7 show data on drug resistance due to the loss of hypoxanthine-guanine phosphoribosyltransferase (HGPRT) enzyme activity with the $r - P_0$ plots. Figure 6.6 includes the data of Morrow (1970) and Fig. 6.7 shows the data of Varshaver et al. (1983). Figures 6.6 and 6.7 demonstrate that the r and P_0 values obtained in this way, are not matched by the $r - P_0$ plots of the Luria–Delbrück, Galton–Watson, and Markov branching process models (from left to right).

To visualize the extent of separation of data from the single-stage models, we carried out confidence interval (CI) analysis of data. The results are depicted in Figs. 6.3, 6.4, 6.6, and 6.7. The vertical error bars are the exact 0.95 CIs for P_0, based on binomial distribution, and corrected for plating efficiency. It is difficult to carry out exact analysis for r because its distribution is complicated. Therefore, we only plotted horizontal bars, the right ends of which correspond to the upper 0.95 quantile of the sample. This analysis shows systematic departures from the single-stage model.

6.1.7 The Two-Stage Model Versus Data

Kimmel and Axelrod (1994) demonstrated that the two-stage model better explains the experimental data concerning drug resistance. The typical estimates of the first-stage forward mutation rate are $\alpha_{01} \approx 10^{-6}$. The corresponding reversal

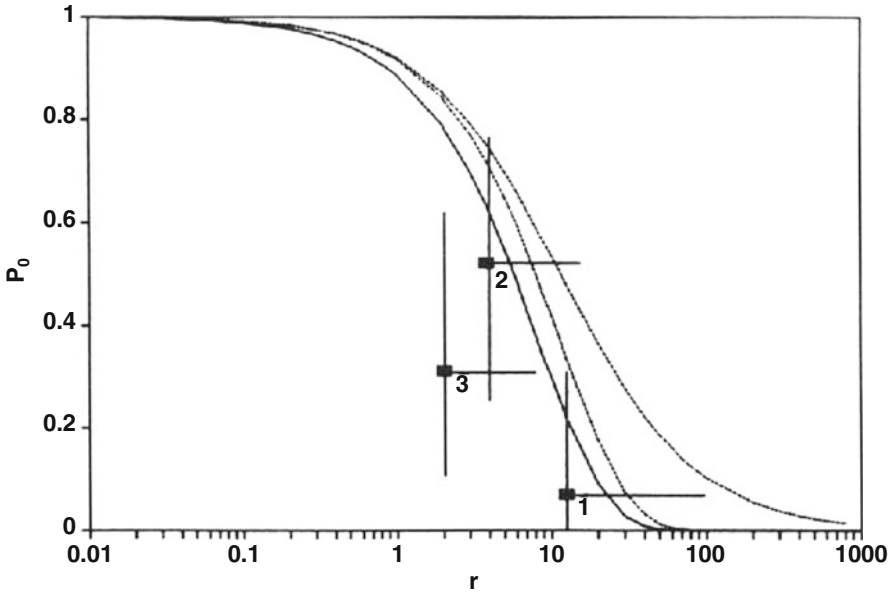


Fig. 6.7 The drug resistance data for Chinese hamster $237-4$ cells from replicate cultures 1, 2, and 3, HPRT$^-$ (Varshaver et al. 1983), and the $r - P_0$ plots ($N = 10^5$) of the Galton–Watson, Luria-Delbrück and Markov branching process models (from left to right). (Source: Kimmel and Axelrod 1994)

rates are $\alpha_{10} \approx 0.2 - 0.95$. Finally, the second-step forward mutation rates are $\alpha_{12} \approx 0.01 - 0.15$. Detailed explanations and a discussion can be found in the original paper. However, let us note that the use of a two-stage model is justified only after the possibilities of fitting the data using the simpler single-stage models were exhausted.

Together, these results suggest that some cases of drug resistance do not result from a single irreversible mutation, but may result from two mutations, the first of which is reversible.

6.1.8 Modified Median Estimator of Mutation Rates

The widely used estimators of mutation rates are: the P_0 estimator (Luria and Delbrück 1943), the median estimator (Lea and Coulson 1949), the Lea–Coulson estimator (1949), and the maximum likelihood estimator (MLE; Jones et al. 1993; Zheng 2002) as well as Bayesian estimators (Asteris and Sarkar 2002), as reviewed by Foster (2006). However, most of these estimators, such as P_0, mean and median, do not (and cannot) consider variations in population sizes of the parallel cultures, but

assume the same population size for each parallel culture. This assumption improves computational accessibility, but may cause bias in estimation.

A modified median estimator that is robust to unequal population sizes is a generalization of the median estimator of Lea and Coulson (1949). In addition to allowing unequal population sizes, N_t, of the parallel cultures, it helps to reduce the estimation variability. Simulation results show a good accuracy and robustness of the modified median estimator compared with the median estimator and the MLE. Details are provided in Wu et al. (2009), where its properties are discussed and compared to two other estimators: the median estimator and the MLE. The P_0 estimator is not discussed due to its known drawbacks (Zheng 2002). Also, the Luria and Delbrück distribution (LDD) is not discussed since the mean estimator (method of moments) tends to be biased because of its long-tail property.

It is instructive to compare the median estimator, MLE, and modified median estimator.

Median Estimator

A general way to find the median estimator is to equate the median of the distribution to the empirical median number of mutants based on all cultures in the batch,

$$\hat{\mu} : P(K \leq k_0 | N_t, \hat{\mu}) = \frac{1}{2}, \tag{6.34}$$

where k_0 is median number of mutants based on all cultures of the batch.

Numerical calculation of the median estimator relies on the cumulative distribution function (cdf) of the LDD, which involves iterative computation of the probabilities. Therefore, a large value of k_0 (usually for $k_0 > 5000$) will lead to computational problems.

The Lea–Coulson estimator is a good approximation in the case of large k_0 (Lea and Coulson 1949). It satisfies the empirical relation, Eq. (37) of Lea and Coulson (1949):

$$\frac{k_0}{\hat{m}} - \log(\hat{m}) = 1.24, \tag{6.35}$$

where m is the expected number of mutations.

Maximum Likelihood Estimator

The first explicit and practical algorithm for computing the MLE of m was published by Jones et al. (1993). To evaluate the reliability of the MLE, Stewart then provided a systematic method for constructing CIs (Stewart 1994). Inspired by their work, Zheng (1992) perfected the MLE derivation and proposed a computationally feasible method for calculation.

The MLE of m is given by:

$$\hat{m} = \arg\max_m L(m|k_1, \ldots, k_n, N_{t1}, \ldots, N_{tn}), \tag{6.36}$$

where $L(m|k_1, \ldots, k_n, N_{t1}, \ldots, N_{tn}) = \sum_{i=1}^{n} \log P(k_i|m, \phi)$.

From the asymptotic distribution of MLE, Zheng further developed a Wald-type $100(1-\alpha)\%$ interval estimation of mutation rate as

$$\frac{\hat{m}}{N_t} \pm \frac{z_{\alpha/2}}{N_t\sqrt{nI(\hat{m})}}, \tag{6.37}$$

where $I(m)$ is the Fisher information defined by:

$$I(m) = E_K\left[\frac{\partial \log P(K|m, \phi)}{\partial m}\right]^2. \tag{6.38}$$

Similar to the median estimator, the MLE also has computational problems when k_0 is large.

Modified Median Estimator

In general, the previously discussed MLE is elegant and easy to compute, but its CI depends on the asymptotic distribution, which is usually not realizable under experimental conditions. Therefore, it has a disadvantage when the sample size (number of parallel cultures) is not large enough. Furthermore, the MLE uses all the information in the data to find the mode of the likelihood function, and it may not be robust with respect to outliers.

In contrast to the MLE, the median estimator is robust by its nature. However, it only uses partial information carried by the data, i.e., the number of mutants in the median culture of the batch, and discards the information contained in other cultures. Moreover, under experimental conditions, there often exists a serious spread of N_t in the batch of parallel cultures. The number of mutants in the median culture does not necessarily result in median estimate of mutation rate in such circumstances. Wu et al. (2009) developed a generalized version of the median estimator, called the modified median estimator. It is defined in two steps: first, for every culture, calculate the median estimator $\hat{\mu}_i$, and second, calculate the median of these estimators in all cultures,

$$\hat{\mu} = \text{median}(\hat{\mu}_i), \tag{6.39}$$

where $\hat{\mu}_i : P(K_i \leq k_i|N_{ti}, \hat{\mu}_i) = \frac{1}{2}$. Detailed derivations of the modified median estimator are given in Appendix to Wu et al. (2009). The estimator was extensively tested by simulation.

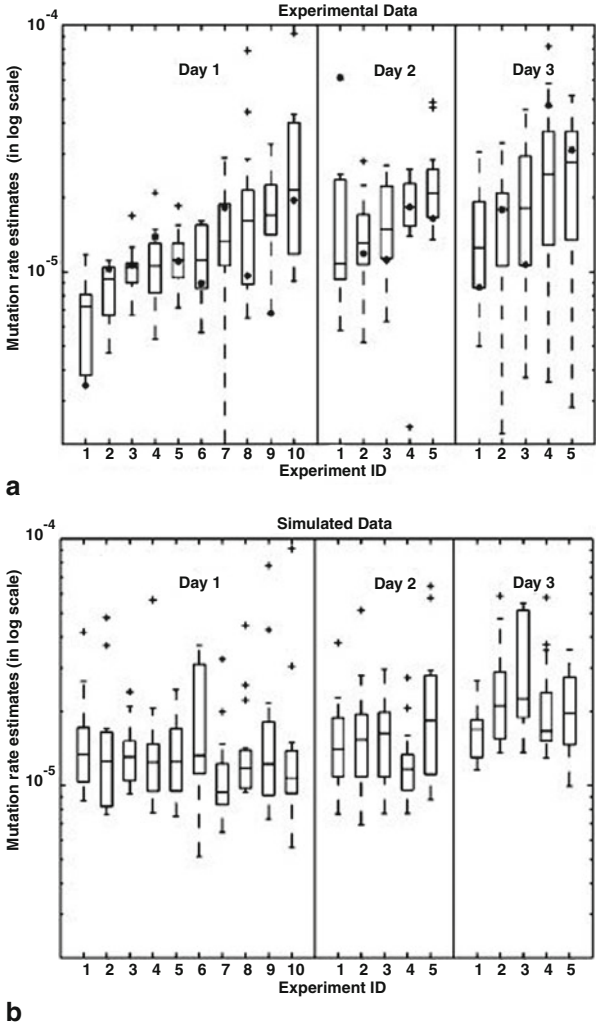


Fig. 6.8 **a** Box plot of mutation rate estimates in 15 replicate experiments in 2 days, ordered by the point estimates. Within each box, the middle line represents the modified median estimator and the asterisk represents the median estimator. **b** Box plot of mutation rate estimates in simulated data, ordered by the point estimates. Each box represents summary statistics of mutation rate estimates in the 15 parallel cultures of each strain. The box has lines at the lower quartile, median, and upper quartile of the data. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. The whiskers indicate the minimum and maximum data values, unless outliers (marked using +) are present, in which case, the whiskers extend to a maximum of 1.5 times the interquartile range. (Source: Wu et al. 2009)

6.1.9 Modified Median Estimator Versus Data

The modified median estimator was chosen to analyze experimental data because of its robustness and computational accessibility (compared to the maximum likelihood method) and favorable coverage and accuracy (compared to the method of median).

Variability of Estimates of Mutation Rates in Yeast

Mutation events in the yeast *Saccharomyces cerevisiae* have been reported as conversion from growth inhibition in the presence of the drug canavanine to growth resistance. Figure 6.8a shows the distribution of mutation estimates (using the method of median for each k, N_t of a culture) coming from 15 parallel cultures assayed in one experiment (Wu et al. 2009). The middle line in the box indicates the mutation rate derived by the modified median estimator, whereas the asterisk indicates the mutation rate derived using the standard median estimator.

Mutation rates estimated from independent replicas of the same experiments would be expected to show limited variability. However, Wu et al. (2008) reported that mutation rates showed variability from experiment to experiment that exceeded expectations. This can be seen by comparison of the experimentally observed data in Fig. 6.8a with the simulated data in Fig. 6.8b, which use the same population sizes and a common mutation rate (combined over all of the independent experiments).

CIs are also helpful in judging the variability of the estimates. CIs of different experiments that do not overlap indicate large variability and poor reproducibility of mutation rate estimates. Comparison of experimental data of Wu et al. 2009, Fig. 6.9a, and simulated data, Fig. 6.9b, indicates that the mutation rates between different independent cultures based on the Luria and Delbrück analysis of experimental data have larger variability than those based on the simulated data. This suggests that the experimental conditions deviate from the Luria and Delbrück assumptions.

Dependence of Estimation of Mutation Rates on Yeast Population Size

One of the assumptions of the Luria and Delbrück model is that mutation rate estimates do not depend on final population size, as shown in the simulated data of Fig. 6.10b. However, Wu et al. (2009) reported that mutation rates estimated from the Luria and Delbrück model were clearly dependent on the final population size, Fig. 6.10a. Under the experimental conditions of these experiments the estimates of mutation rates by the Luria and Delbrück model were inversely dependent on population size ranging from 10^6 to 2×10^7 cells per culture.

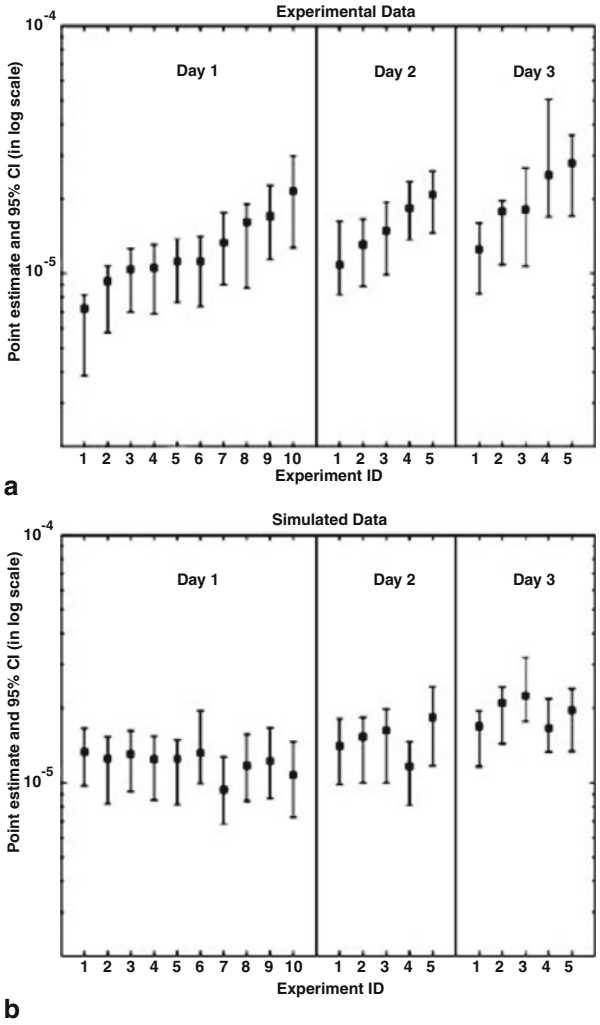


Fig. 6.9 **a** Confidence intervals (*CI*) of mutation rate estimation in 15 replicate experiments in 2 days, ordered by the point estimates. **b** CIs of mutation rate estimates in simulated data, ordered by the point estimates. (Source: Wu et al. 2009)

The differences in the mutation rates estimated by the modified and standard median estimators reflect differences of the final population sizes of the parallel cultures. The modified median estimator reduces the component of variability of estimates introduced by unequal final population sizes. In addition, as demonstrated on simulated data (see Wu et al. 2009 for details), the modified median estimator is robust with respect to outliers in the data and departures from the Luria and Delbrück distribution.

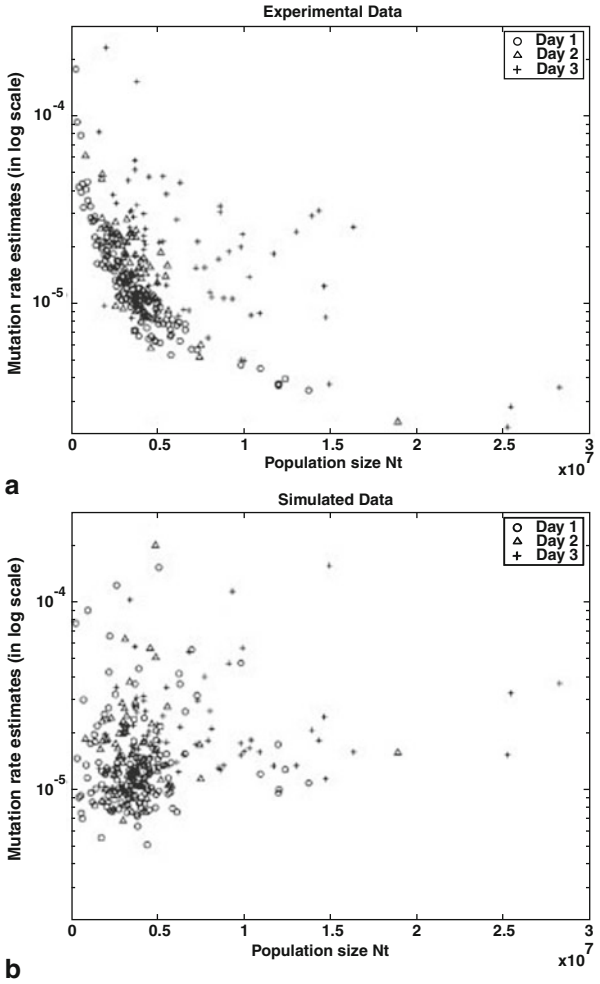


Fig. 6.10 Scatter plot of mutation rate estimates versus population sizes. **a** yeast data, **b** simulated data using constant mutation rate assumption. (Source: Wu et al. 2009)

Independence of Estimated Mutation Rate on Bacterial Population Size

The Luria and Delbrück model was originally used to estimate mutation rates in bacterial cultures. Since the Luria and Delbrück method of analysis of yeast cultures resulted in estimates of mutation rate that depended on population size (see above), the possible effect of population size on the estimation of mutation rates of bacterial cultures needs to be reconsidered. Figure 6.11a shows the estimated mutation rate using the modified median estimator, as a function of bacterial population size (Hastings et al. 2000). There is no dependence on population size. Figure 6.11b shows

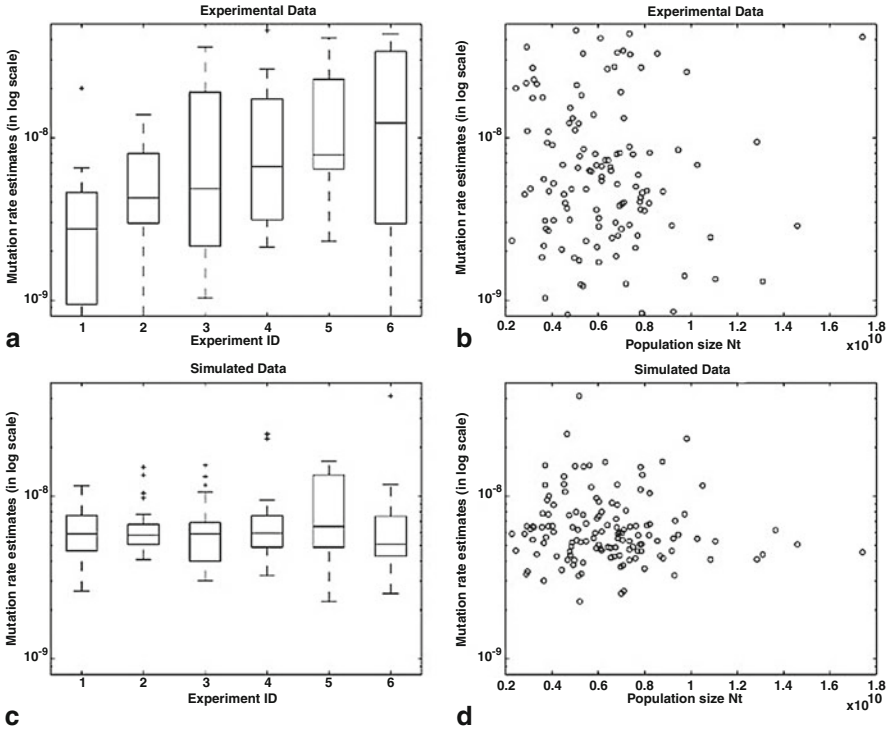


Fig. 6.11 Scatter plot of mutation rate estimates versus population sizes. **a** bacterial data, **b** simulated data using constant mutation rate assumption. (Source: Wu et al. 2009)

the simulated data assuming independence of estimated mutation rate on population size. The similarity indicates that the Luria and Delbrück model describes well the mutations in the bacterial data.

Estimates of mutation rates in yeast cultures show a dependence on final population size, but estimations of mutation rates in bacterial cultures do not show a dependence on final population size. This difference may be due to the 10^3-fold difference in the range of final population sizes in the two cultures. Or, it may be due to the difference mechanism of mutation to drug resistance in the two cultures, single base change in bacteria, and chromosome loss or mitotic recombination in yeast.

6.1.10 Recent Developments in Theory and Application of Fluctuation Analysis

We will focus discussion on recent works, which attempt to explore applicability of fluctuation analysis (FA) under modified mathematical assumptions, suitable for new

experimental approaches. Somewhat idiosyncratically, we will discuss mathematical models by Angerer (2001), estimation approaches by Zheng (1999, 2002, 2005) and applications to genomic mutation rates in bacteria by Foster (2006).

Angerer (2001) developed a theory of FA which relaxed some of the Luria–Delbrück hypotheses regarding proliferation and mutation models. To estimate the mutation rate, one requires a mathematical model of the Luria–Delbrück distribution (LLD) of the number of mutants that a culture contains just before a sample is removed to determine the number of mutant cells and the total number of cells in the culture. Angerer presents such a model assuming a Bellman–Harris process of cell proliferation, and shows that under natural assumptions concerning the lifetime distributions and the offspring distributions of mutant and nonmutant cells, the suitably normed and centered number of mutants contained in a large culture of proliferating cells converges to a stable random variable with index 1. This result is obtained under the assumption that the mutation under consideration is *neutral* in the sense that on average and in the long run, mutant cells and nonmutant cells produce offspring at the same rate. This work addresses one aspect of the FA, which has been causing unease among biologists: the reliance on assumptions of either exponentially distributed or constant lifetimes of cells. In the Bellman–Harris branching process, these distributions can be essentially arbitrary, but explicit solutions are not expected. Therefore, Angerer develops asymptotic approximations.

Another essential problem has been tackled by Zheng (2005). When inferring the mutation rate from an experiment, it is frequently assumed that the number of mutants in each test tube follows a common distribution. This assumption conceptually restricts the scope and applicability of fluctuation assay. This assumption is relaxed by Zheng by proposing a Bayesian two-level model, under which an experiment-wide average mutation rate can be defined. The new model suggests a gamma mixture of the Luria–Delbrück distributions. The mixture model also offers a practical Markov chain Monte Carlo method for estimating mutation rates. The work has been motivated by a concern of mutation researchers about assumptions underpinning the fluctuation assay. The problem can be explained using the example of a simplified bacterial fluctuation experiment. Suppose there are n test tubes that contain a liquid culture, with n usually ranging from 10 to 100. Each tube is seeded with a small number (denoted N_0) of wild-type cells sensitive to the selective agent. At the end of the incubation period, the total number of sensitive wild-type cells per tube is estimated by diluting the culture and determining the number of single-cell-derived colonies that form on solid medium without the selective agent. The number of selective-agent-resistant mutant cells is estimated by diluting the culture and determining the number of single-cell-derived colonies that form on solid medium with the selective agent. In the classical model (Lea and Coulson 1949), growth of nonmutant cells is assumed to be deterministic and identical in all tubes. However, in the laboratory the final number of cells $N_T(i)$ can vary considerably across tubes in the same experiment, which violates a basic assumption underlying the current statistical procedure for estimating mutation rates from fluctuation assay data. To address this problem, Wu et al. (2009), discussed in Sect. 6.1.8, offered a robust estimator based on the median number of mutants of the n cultures. Their method requires the measurement

of individual $N_T(i)$ for all cultures. This requirement may present a problem because in practice biologists often sample only a small number (e.g., three) of cultures to obtain an average of N_T.

To offer a practical alternative, Zheng (2005) assumes that tubes in an experiment inherently have slightly different values of cell growth rate m, which is reflected by a large variation in N_T. As one cannot accurately estimate the actual m for each, Zheng proposes averaging the values of m via a two-level Bayesian model. Specifically, he assumes that the ith tube in the experiment has m_i as its value of the parameter m. These m_i are considered as drawn from a common parental distribution.

Foster's group (Lee et al. 2012) addressed the rate and molecular spectrum of spontaneous mutations in the bacterium *E. coli* as determined by whole-genome sequencing. They analyzed spontaneous mutations accumulated for more than thousands of generations by wild-type *E. coli* and a derivative defective in mismatch repair (MMR) of DNA, the primary pathway for correcting replication errors. Among their conclusions are (i) the mutation rate of a wild-type *E. coli* strain is $\sim 1 \times 10^{-3}$ per genome per generation; (ii) mutations in the wild-type strain have the expected mutational bias for $G:C > A:T$ mutations, but the bias changes to $A:T > G:C$ mutations in the absence of MMR; and (iii) although the rate of small ($\leq$4 nucleotides) insertions and deletions is high at repeat sequences, these events occur at only one tenth the genomic rate of base-pair substitutions. Bacteria isolated from nature often lack MMR capacity, suggesting that modulation of MMR can be adaptive, this latter being an important finding.

The Foster group has been pioneering techniques related to fluctuation analysis in order to better understand the contribution of mutation to evolution. They have developed methods to improve the accuracy of determining mutation rates and to characterize the molecular spectrum of mutations. In addition, their focus has included the question as to whether, and how, intrinsic and extrinsic factors influence mutational processes. Their present approach involves the mutation-accumulation (MA) strategy designed to allow mutations to occur in a neutral manner, devoid of selective pressure. The general strategy is to establish a number of clonal populations from a founder individual and then to take each population through repeated single-individual bottlenecks for thousands of generations. Because it can be assumed that the approximate effective population is one, genetic drift prevents selection from eliminating all but the most deleterious mutations, which typically are less than 1 % of mutations. The experiments in Lee et al. (2012) were designed to yield a highly accurate estimate of the spontaneous mutation rate of *E. coli*. The remarkably low rates at which mutations occur are interpreted as resulting from both the high intrinsic accuracy of DNA replication and various enzymatic activities that survey and repair DNA. In addition, for comparison they provide (Table 2 in Lee et al. 2012) mutation rates for the wild-type strain estimated using classical fluctuation tests, which are 6–9 six to nine times *lower* than those resulting from MA. The explanation provided involves phenotypic delay following mutation events in FA. However, a mathematical theory of MA strategy has not been developed, so the biases of that method are not known.

Niccum et al. (2012) present a method of estimation of mutation rates in growing cell colonies which is not fluctuation analytic in the strict sense of the term. However, it is based on a branching process model. The authors present a rigorous mathematical solution to the mutation rate problem using an unbiased and consistent estimator. Using this estimator they demonstrated, based on the properties of the Galton–Watson process, that mutation rates can be calculated by determining mutant accumulation, that is, from the number of mutants measured in two successive generations. Consistency of the estimator is verified by simulations. These show a rapid convergence to the targeted mutation rate which is reached between the 25th and 30th generations.

6.2 The Positive Regular Case of the Multitype Galton–Watson Process

We study in this section the variant of the multitype Galton–Watson process, the behavior of which is a direct extension of the single-type case. We proceed as in Chap. 2 of Harris (1963). In the previous section, we used some of these results based on intuitive generalizations. An authoritative and advanced source on multitype classical processes is provided in the book by Mode (1971), which can be used as a reference for most of this chapter.

We follow evolution of a population composed of particles of k types. An ancestral particle of type i lives for a unit time interval and in the moment of death produces a random number of progeny particles of generally all k types. The numbers of its progeny of all types constitute a random vector with nonnegative integer entries, characterized by pgf $f^i(s_1, \ldots, s_k)$. A progeny of type j starts, independently of all other progeny, a subprocess with itself as the ancestor, by producing at the moment of death, a random vector of progeny of all types, characterized by pgf $f^j(s_1, \ldots, s_k)$. The distribution of this subprocess depends only on the type of the ancestral particle.

The counts of particles of all types existing at time n in the process started by an ancestor of a fixed type constitute a random vector denoted $\mathbf{Z}_n = (Z_n^1, \ldots, Z_n^k)$. The distribution of this vector depends on the type of the ancestral particle of the process. Below, we provide a definition and a theorem stating that in the multitype process, the pgf's of $\mathbf{Z}_n$ are functional iterates of the progeny pgf's, as it was the case for the single-type Galton–Watson process.

6.2.1 Basics

The following definition uses a forward approach to the process by relating the numbers of particles in generation $n + 1$ to those in the preceding generation n. In this way, it underscores the Markov character of the multitype Galton–Watson process.

Definition 6.1 *Let T denote the set of all k-dimensional vectors whose components are nonnegative integers. Let $\mathbf{e}_i, 1 \le i \le k$, denote the vector whose ith component is 1 and whose other components are 0.*

The multitype (or vector) Galton–Watson process *is a temporally homogeneous vector-valued Markov process $\mathbf{Z}_0, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$, whose states are vectors in T. We shall always assume that $\mathbf{Z}_0$ is nonrandom. We interpret Z_n^i the ith component of $\mathbf{Z}_n$, as the number of objects of type i in the nth generation.*

The transition law for the process is as follows. If $\mathbf{Z}_0 = \mathbf{e}_i$, then $\mathbf{Z}_1$ will have the generating function

$$f^i(s_1, \ldots, s_k) = \sum_{r_1, \ldots, r_k}^{\infty} p^i(r_1, \ldots, r_k) s_1^{r_1}, \ldots, s_k^{r_k}, \ |s_1|, \ldots, |s_k| \le 1, \tag{6.40}$$

where $p^i(r_1, \ldots, r_k)$ is the probability that an object of type i has r_1 children of type 1, $\ldots$, r_k of type k. In general, if $\mathbf{Z}_n = (r_1, \ldots, r_k) \in T$, then $\mathbf{Z}_{n+1}$ is the sum of $r_1 + \cdots + r_k$ independent random vectors, r_1 having the generating function f^1, r_2 having the generating function f^2, $\ldots$, r_k having the generating function f^k. If $\mathbf{Z}_n = 0$, then $\mathbf{Z}_{n+1} = 0$.

The generating function of $\mathbf{Z}_n$, when $\mathbf{Z}_0 = \mathbf{e}_i$, will be denoted by $f_n^i(s_1, \ldots, s_k) = f_n^i(\mathbf{s})$ $i = 1, \ldots, k$ $n = 0, 1, \ldots$. Then f_1^i is the function f^i of Eq. (6.40). The vector $(f_n^1(\mathbf{s}), \ldots, f_n^k(\mathbf{s}))$ will be frequently denoted $\mathbf{f}_n(\mathbf{s})$.

Directly from this definition, we can deduce the following theorem: We omit the details as they are an extension of those in Sect. 3.1.2. They can be obtained by a direct application of Theorem A.1, part 6.

Theorem 6.1 *The generating functions f_n^i are functional iterates, defined by the relations*

$$f_{n+1}^i(\mathbf{s}) = f^i[f_n^1(\mathbf{s}), \ldots, f_n^k(\mathbf{s})], \ n = 0, 1, \ldots ;$$
$$f_n^0(\mathbf{s}) = s_i, \ i = 1, 2, \ldots, k. \tag{6.41}$$

More generally, we have, in vector form

$$\mathbf{f}_{n+N}(\mathbf{s}) = \mathbf{f}_n[\mathbf{f}_N(\mathbf{s})], \ n, N = 0, 1, 2, \ldots . \tag{6.42}$$

We define $\mathbf{M} = (m_{ij})$ to be the matrix of expected numbers of progeny of all types of parent particles of all types. Specifically, $m_{ij} = \mathrm{E}(Z_1^j | \mathbf{Z}_0 = \mathbf{e}_i) = \frac{\partial f^i(1, \ldots, 1)}{\partial s_j}$, $i, j = 1, \ldots, k$ is the expected number of progeny of type j of a particle of type i. It is assumed that all the first moments m_{ij} are finite and not all equal to 0. By using chain rule in (6.41), we obtain $\mathrm{E}(\mathbf{Z}_{n+1} | \mathbf{Z}_n) = \mathbf{Z}_n \mathbf{M}$. More generally,

$$\mathrm{E}(\mathbf{Z}_{n+N} | \mathbf{Z}_N) = \mathbf{Z}_N \mathbf{M}^n. \tag{6.43}$$

Analogous expressions for variances are more complicated (see Harris 1963; Mode 1971).

6.2.2 Positivity Properties

The following are the essentials of the Perron–Frobenius theory of positive matrices: This theory demonstrates that iterates of positively regular nonnegative matrices can be approximated using the powers of the dominating eigenvalue of the matrix, which is shown to be positive. As a consequence, the asymptotic properties of the multitype Galton–Watson process in the positive regular case can be expressed using powers of this eigenvelue.

We shall call a vector or a matrix *positive, nonnegative, or* 0 if all its components have the following properties. If $\mathbf{u}$ and $\mathbf{v}$ are vectors or matrices, then $\mathbf{u} > \mathbf{v}$ ($\mathbf{u} \geq \mathbf{v}$) means that $\mathbf{u} - \mathbf{v}$ is positive (nonnegative). Absolute value signs enclosing a vector or a matrix denote the sum of the absolute values of the elements, e.g., $|\mathbf{Z}_n| = \sum_i |Z_n^i|$.

Theorem 6.2 *Let* $\mathbf{M}$ *be a nonnegative matrix of order k, which is irreducible, i.e., such that* $\mathbf{M}^N$ *is positive for some positive integer N. Then* $\mathbf{M}$ *has a positive eigenvalue* ρ *that is simple and greater in absolute value than any other eigenvalue;* ρ *corresponds to positive right and left eigenvectors* $\mu = (\mu^i)$ *and* $\nu = (\nu^i)$*, which are the only nonnegative eigenvectors. Moreover, we have*

$$\mathbf{M}^n = \rho^n \mathbf{M}_1 + \mathbf{M}_2^n, \; n = 1, 2, \ldots, \tag{6.44}$$

where $\mathbf{M}_1 = (\mu^i \nu^j)$*, with the normalization* $\sum_i \mu^i \nu^i = 1$*. Hence* $\mathbf{M}_1 \mathbf{M}_1 = \mathbf{M}_1$*. Furthermore,*

1. $\mathbf{M}_1 \mathbf{M}_2 = \mathbf{M}_2 \mathbf{M}_1 = 0$.
2. $|\mathbf{M}_2^n| = O(\alpha^n)$ *for some* $\alpha \in (0, \rho)$.
3. *If j is a positive integer then* ρ^j *corresponds to* $\mathbf{M}^j$ *in the same manner as* ρ *corresponds to* $\mathbf{M}$.

A multitype Galton–Watson process is called *positively regular or irreducible* if $\mathbf{M}^n$ is positive for some positive integer N.

6.2.3 Asymptotic Behavior in the Supercritical Case

The following result is a direct extension of the analogous result for the single-type process (Theorems 3.4 and 3.5):

Theorem 6.3 *Suppose the process is positively regular with* $\rho > 1$*, and all the second moments of progeny distributions are finite. Then the random vectors* $\mathbf{Z}_n / \rho^n$ *converge with probability 1 to a random vector* $\mathbf{W}$*. Vector* $\mathbf{W}$ *is nonzero except for trivial cases of all variances* $\mathbf{V}_i$ *being zero or* $\mathbf{Z}_0 = \mathbf{0}$*. If* $\mathbf{W}$ *is nonzero, then with probability 1 its direction coincides with that of* ν*, the left eigenvector of* $\mathbf{M}$*.*

One of the consequences of the theorem is that the limit law in the positively regular case is strictly one-dimensional. While the total number of particles is subject to wide dispersion, their proportions become constant with probability 1.

6.2.4 Probability of Extinction

It is understandable that the probability of extinction of a multitype process depends on the type of its ancestral particle. Otherwise, the rule is analogous as in the single-type case (Sect. 3.3). Let q^i be the extinction probability, if initially there is one object of type $i = 1, 2, \ldots, k$. That is $q^i = \mathrm{P}\{\mathbf{Z}_n = 0 \text{ for some } n | \mathbf{Z}_0 = \mathbf{e}_i\}$. The vector $(q^1, \ldots, q^k)$ is denoted by $\mathbf{q}$.

Theorem 6.4 *Suppose the process is positively regular and not singular (which would mean that each object has exactly one progeny). If* $\rho \leq 1$, *then* $\mathbf{q} = \mathbf{1}$. *If* $\rho > 1$, *then* $\mathbf{0} \leq \mathbf{q} < \mathbf{1}$ *and* $\mathbf{q}$ *satisfies the equation*

$$\mathbf{q} = \mathbf{f}(\mathbf{q}). \tag{6.45}$$

6.3 Application: A Model of Two-Cell Populations

The example we present is a simplified version of the model considered in Kimmel and Arino (1991). It is motivated by an experiment described in Sennerstam and Strömberg (1984).

Let us consider two cell populations evolving according to the following rules:

1. Both populations have fixed interdivision times equal to 1.
2. In both populations the divisions are entirely effective, i.e., each parent cell produces exactly two progeny initially of the same type.
3. After division, each type 1 progeny (independent of the other) switches to type 2 with probability p_{12}, and remains type 1 with probability $p_{11} = 1 - p_{12}$.
4. Analogously, each type 2 progeny (independent of the other) switches to type 1 with probability p_{21}, and remains type 2 with probability $p_{22} = 1 - p_{21}$.

The known biological example is the population of cultured transformed embryonic cells maintained by Sennerstam. The "normal" embryonic cell has a program to switch irreversibly from one developmental stage to the next. The transformed cells are maintained indefinitely since they switch back and forth between two stages, named by us 1 and 2. Under the simplified assumptions specified above, their proliferation is described by a 2-*type Galton–Watson process*.

The progeny pgf's of the process are

$$f^1(s_1, s_2) = (p_{11}s_1 + p_{12}s_2)^2, \tag{6.46}$$

$$f^2(s_1, s_2) = (p_{21}s_1 + p_{22}s_2)^2. \tag{6.47}$$

The expected progeny matrix of the process is equal to

$$\mathbf{M} = \begin{pmatrix} 2p_{11} & 2p_{12} \\ 2p_{21} & 2p_{22} \end{pmatrix}. \tag{6.48}$$

The eigen values of matrix $\mathbf{M}$ are found from the equation

$$\rho^2 - \rho(2p_{11} + 2p_{22}) + 4(p_{11}p_{22} - p_{12}p_{21}) = 0.$$

The greater of the two real roots of this equation (the Perron–Frobenius root or eigenvalue) is equal to $\rho = 2$. The left eigenvector ν corresponding to the Perron–Frobenius root is the row vector satisfying the matrix equation $\nu(\mathbf{M} - 2\mathbf{Id}) = \mathbf{0}$, or

$$2(\nu_1, \nu_2) \begin{pmatrix} -p_{12} & p_{12} \\ p_{21} & -p_{21} \end{pmatrix} = \mathbf{0}. \tag{6.49}$$

We obtain

$$\frac{\nu_1}{\nu_2} = \frac{p_{21}}{p_{12}}.$$

The process is positively regular. Theorem 6.3 yields that with probability 1

$$(Z_n^1, Z_n^2) \sim 2^n(\nu_1, \nu_2)W, \quad n \to \infty,$$

where W is a scalar random variable.

The meaning of this result is that the proportion of the type 1 and type 2 cells is asymptotically determined by the ratio $\frac{\nu_1}{\nu_2} = \frac{p_{21}}{p_{12}}$. The interesting feature is that both p_{21} and p_{12} can be very small, i.e., that the switching between both types is not frequent, and still the proportion is maintained. For the experimental data, it was estimated that p_{12} and p_{21} are of the order of 0.1 (Arino and Kimmel 1991).

6.4 Application: Stochastic Model of the Cell Cycle with Chemotherapy

The current application does not draw on the theory in the previous section. Instead, it is an example of a model using a multitype Bellman–Harris process.

The goal of cancer chemotherapy is to stop tumor cells from dividing and to kill them while sparing normal cells. Some chemotherapy protocols depend on the differential effect of drugs on cells in different compartments of the cell cycle. For instance, combination drug chemotherapy may utilize two drugs which affect cells in different compartments of the cell cycle with different efficiencies. Such combination chemotherapy is expected to be more effective in tumor cell populations than in normal cell populations. The rationale is that tumor cell populations have a larger fraction of cells progressing through the cell cycle than normal cells. This approach requires knowledge of the "drug action curve," the percentage of cells affected depending on their position in the cell cycle. In Sect. 5.4, we developed a method of

estimating the duration of cell cycle compartments, based on stathmokinetic experiments. This method is now extended to determine the relative effects of a drug on cells in different compartments of the cell cycle.

Modern technology allows determining the amount of DNA per cell by measuring the fluorescence of stained DNA excited by a laser in a flow cytometer. This has lead to an improved stathmokinetic method that utilizes the amount of DNA per cell, rather than the number of cells in mitosis, as a function of time for which the cells are exposed to the stathmokinetic agent. The means and variances of durations of each of the cell cycle compartments can then be estimated using the mathematical methods described in Sect. 5.4.

Additional mathematical methods are required to obtain the estimates of the cell cycle specific effects of anticancer drugs. We develop a model, which describes the flow of cells through successive compartments of the cell cycle. The model allows estimation of the fraction of cells blocked in each cell cycle compartment by an anticancer agent.

This application is mainly based on the paper by Kimmel and Traganos (1986). It is the continuation of the stathmokinetic analysis example of Sect. 5.4. The mathematical tool we use is the multitype Bellman–Harris process. We do not develop a rigorous theory, but employ intuition and analogies with the Galton–Watson branching process.

We want to model the long-term in vitro effects of an anticancer drug, acting with a different strength on cells in different phases of the cell cycle, based on the short-term observations collected using the stathmokinetic experiment. For this purpose, we decompose the cell cycle into a sequence of compartments differing with respect to sensitivity to the drug. These compartments may be different from individual cell cycle phases. Specifically, in the current model, the S phase is subdivided into a number of smaller compartments, to account for different sensitivities of cells in different stages of DNA synthesis.

6.4.1 Model of Drug-Perturbed Stathmokinesis

The following model is employed to analyze the drug action (Fig. 6.12): The cell cycle is divided into M disjoint compartments. Cell residence time in the mth compartment is an independent random variable with distribution density $p_m(\cdot)$. The conditions of the stathmokinetic experiment are satisfied, by assuming that there is no cell inflow into the first compartment nor cell outflow from the last (Mth) compartment. In each compartment, exposure to a given concentration of the drug causes a permanent block for a fraction $1 - u_m$ of cells, which would otherwise leave this compartment. By choosing a sufficiently dense subdivision of the cell cycle into compartments, it is possible to construct a curve of drug action, the coordinates of which are the quantities $1 - u_m$.

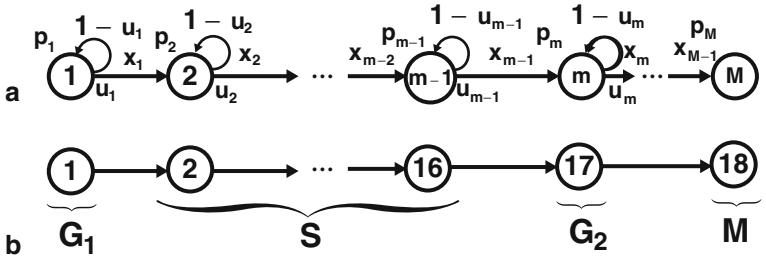


Fig. 6.12 **a** The model of blocking drug action. The cell cycle is divided into M compartments. There is no cell flow into the first compartment, nor cell outflow from the last compartment. Notation: $p_m(t)$, distribution density of the residence time; $x_m(t)$, outflow rate; $N_m(t)$, cell count: u_m, cell fraction in the mth compartment not blocked by the drug. **b** Correspondence between compartment number and cell cycle phase. (Source: Kimmel and Traganos 1986)

Let us denote by $N_m(t)$ the expected cell count in the mth compartment and by $x_m(t)$ the expected cell outflow rate from the mth compartment, at time t. We have

$$\begin{aligned} N_1(t) &= N_1(0) - \int_0^t x_1(s)ds. \\ N_m(t) &= N_m(0) + \int_0^t [x_{m-1}(s) - x_m(s)]ds, \quad m = 2, \ldots, M-1. \\ N_M(t) &= N_M(0) + \int_0^t x_{M-1}(s)ds. \end{aligned} \tag{6.50}$$

It is assumed now that before the beginning of stathmokinesis, i.e., for $t < 0$, the cell population was in the exponential steady state (ESS), i.e., expected cell counts in all the cell cycle compartments were proportional to e^{bt}. The constant b is the Malthusian parameter of exponential growth.

Balancing of expected ESS cell flows from one cell cycle compartment to another, as described in more detail in Kimmel (1980a, b), we obtain

$$\begin{aligned} N_1(0) &= 2(1 - \hat{p}_1), \\ N_m(0) &= 2\hat{p}_1 \cdots \hat{p}_{m-1}(1 - \hat{p}_m) \end{aligned} \tag{6.51}$$

where $\hat{p}_m$ is the Laplace transform of the distribution $p_m(\cdot)$, evaluated at b:

$$\hat{p}_m = \int_0^\infty p_m(t)e^{-bt}dt. \tag{6.52}$$

Computation of the outflows $x_m(\cdot)$ perturbed by the drug is more complicated. Except for $x_1(\cdot)$, the cell outflow is the sum of a component from the outflow of the preceding compartment and another component from the initial distribution (at $t = 0$) of cells in this compartment:

$$x_m(t) = u_m \left[x_{m-1}(t) * p_m(t) + x_m^0(t)\right], \quad m = 2, \ldots, M-1, \tag{6.53}$$

where the asterisk denotes the convolution of functions $(f * g)(t) = \int_0^t f(t-\tau)g(\tau)d\tau$. The flow $x_m^0(t)$ can be calculated in the following way: let us denote by $p_{1\,m}(t)$ the distribution of the sum of residence times in compartments 1 through m, and by $P_{1\,m}(t)$ the corresponding cumulative distribution. Also, let us denote

$$a_{1\,m}(t) = \mathrm{e}^{bt}\int_t^\infty p_{1\,m}(s)\mathrm{e}^{-bs}ds = \mathrm{e}^{bt}\hat{p}_{1\,m} - \mathrm{e}^{bt} * p_{1\,m}(t), \tag{6.54}$$

$$a_m(t) = \mathrm{e}^{bt}\int_t^\infty p_m(s)\mathrm{e}^{-bs}ds = \mathrm{e}^{bt}\hat{p}_m - \mathrm{e}^{bt} * p_m(t). \tag{6.55}$$

We have $\hat{p}_{1\,m} = a_{1\,m}(0)$ and $\hat{p}_m = a_m(0)$. In Proposition 5.1, we found the asymptotics of the number of cells in phase 1 of the cell cycle, when the cell cycle is subdivided into two phases, under normal conditions in a stathmokinetic experiment not perturbed by any other agent. We can consider our compartments 1 through m as a phase 1, and by doing so, we obtain by Proposition 5.1

$$\bar{N}_{1\,m}(t) = 2\left[1 - P_{1\,m}(t) - a_{1\,m}(t)\right]. \tag{6.56}$$

Let us note that, by Eq. (6.54), we have $d[a_{1\,m}(t)]/dt = ba_{1\,m}(t) - p_{1\,m}(t)$, which implies

$$\frac{\mathrm{d}[\bar{N}_{1\,m}(t)]}{dt} = -2ba_{1\,m}(t). \tag{6.57}$$

The outflow $x_m^0(t)$ from the initial distribution of cells in compartment m is the same whether or not a perturbing agent (other than the stathmokinetic agent) is applied. It is equal to the total outflow from compartments 1 through m, minus a component due to the outflow from compartments 1 through $m-1$:

$$\begin{aligned}
x_m^0(t) &= \tfrac{\mathrm{d}[-\bar{N}_{1\,m}(t)]}{dt} - \tfrac{\mathrm{d}[-\bar{N}_{1,m-1}(t)]}{dt} * p_m(t) \\
&= 2b[a_{1\,m}(t) - a_{1,m-1}(t) * p_m(t)] \\
&= 2b\{\mathrm{e}^{bt}\hat{p}_{1,m} - \mathrm{e}^{bt} * p_{1\,m}(t)] \\
&\quad -[\mathrm{e}^{bt}\hat{p}_{1,m-1} - \mathrm{e}^{bt} * p_{1,m-1}(t)] * p_m(t)\} \\
&= 2b\{\mathrm{e}^{bt}\hat{p}_{1,m-1}\hat{p}_m - \mathrm{e}^{bt} * p_{1\,m}(t)] \\
&\quad -[\mathrm{e}^{bt} * p_m(t)\hat{p}_{1,m-1} - \mathrm{e}^{bt} * p_{1,m-1}(t) * p_m(t)]\} \\
&= 2b\hat{p}_{1,m-1}a_m(t).
\end{aligned} \tag{6.58}$$

Combining (6.53) and (6.58), we write down the following recurrence:

$$\begin{aligned}
x_1(t) &= 2ba_1(t), \\
x_m(t) &= u_m[x_{m-1}(t) * p_m(t) + 2b\hat{p}_{1,m-1}a_m(t)], \quad m = 2, \ldots, M-1.
\end{aligned} \tag{6.59}$$

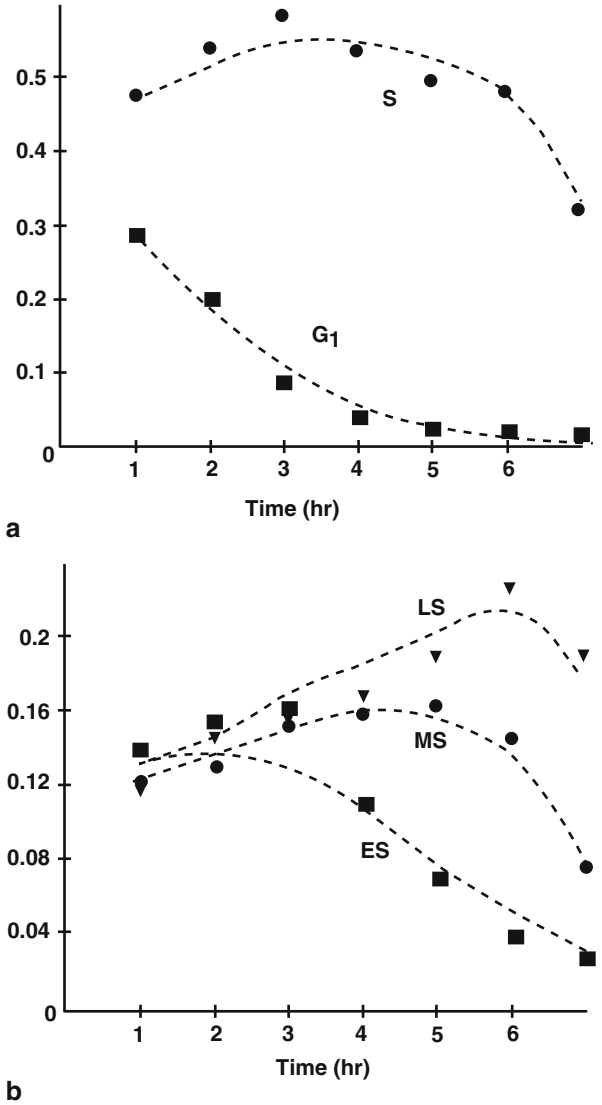


Fig. 6.13 Stathmokinetic data (low drug concentration) fitted by the model curves: (**a**) S (circles), and G_1 (squares) phase, (**b**) early (channels 27–30, squares), mid (channels 32–35, circles), and late (channels 37–40, triangles) S phase "windows." (Source: Kimmel and Traganos 1986)

Based on (6.59) an explicit expression is derived:

$$x_m(t) = 2b\left\{\sum_{i=1}^{m}\left(\prod_{j=1}^{i-1}\hat{p}_j\right)\left(\prod_{j=i}^{m}u_j\right)a_i(t) * [p_{i+1}(t) * p_{i+2}(t) * \cdots * p_m(t)]\right\}, m = 1, \ldots, M-1. \tag{6.60}$$

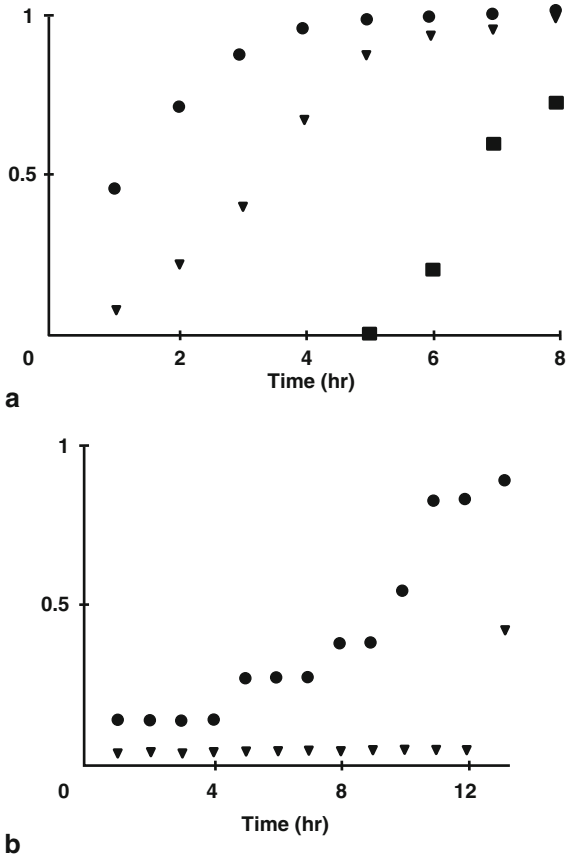


Fig. 6.14 Example of nonparametric estimation of distribution $P_1(t)$ based on exit curve $f_1(t)$.(**a**) Friend erytholeukemia cells: circles, $G_{1\ A}$; triangles, G_1; squares, $G_1 + S$. (**b**) L -cells: circles, G_1; triangles, $G_1 + S$. (Source: Kimmel and Traganos 1986)

6.4.2 Model Parameters

It is generally true that the structure of a model depends on the precision of the measurements. In the present case, we divide the cell cycle into smallest compartments in which it is possible to follow the cell count. A fine subdivision is possible in S phase: We can consider the cells ascending from lower to higher DNA content. Therefore, the model has structure as depicted in Fig. 6.12b: compartment 1 is the G_1 phase, compartments 2–16 cover the S phase, compartment 17 is G_2, and compartment 18 is M.

The main source of variability in the cells' generation time is its transit through the G_1 phase. In practice, the durations of all the other cell cycles phases can be considered nonrandom. The distribution of cell residence time in G_1 was estimated (Fig. 6.14) with the aid of a nonparametric procedure presented in Sect. 5.4.

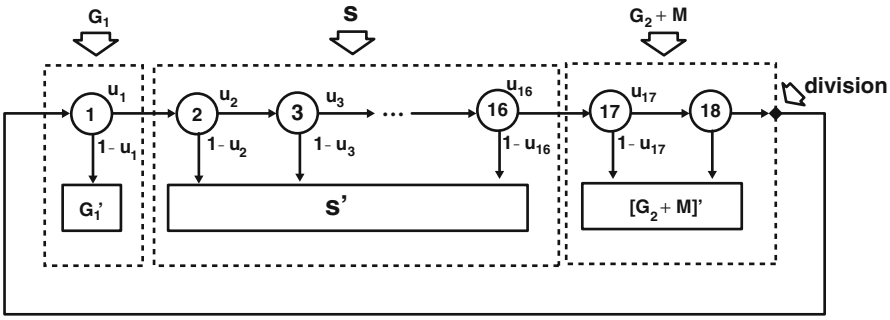


Fig. 6.15 Modelling continuous exposure to the drug. Numbering of the basic model compartments is the same as in Fig. 6.9. It is assumed that cells blocked in given phase do not progress further. Instead, they are trapped in the additional "primed" compartments [G_1', S', and $(G_2 + M)'$]. Compartment $(G_2 + M)'$ includes some of the blocked G_2 cells that had progressed to M before they underwent ineffective division which increased their ploidy. (Source: Kimmel and Traganos 1986)

The deterministic residence times in the remaining cell cycle compartments were assessed based on the ESS cell counts in these compartments. Their estimation as well as estimation of the coefficients u_m characterizing drug action is described in Kimmel and Traganos (1986).

6.4.3 *Prediction of the Effects of Continuous Exposure to the Drug*

Figure 6.15 presents the model used to predict effects of continuous exposure to the drug. It is assumed that once a cell is blocked it does not progress further through the cell cycle. In the model, the blocked cells pass to the "primed" compartments: G_1', S', or $(G_2 + M)'$. The $(G_2 + M)'$ compartment also contains those G_2 cells which progressed to M but did not divide; instead, they increased their ploidy by, for example, defective cytokinesis.

Simulation of the effects of continuous exposure to the drug based on this model was carried out analogously to similar simulations in Kimmel and Traganos (1985) or in Darzynkiewicz et al. (1984); for a more theoretical treatment, see Kimmel (1980c).

6.4.4 *Results*

The estimates of the basic parameters of the cell cycle of exponentially growing Friend erythroleukemia cells are as follows: the average residence time in G_1, $\mathrm{E}(T_{G_1}) = 3.43$ h; the residence times in S, $T_S = 5.08$ h; in G_2, $T_{G_2} = 2.21$ h;

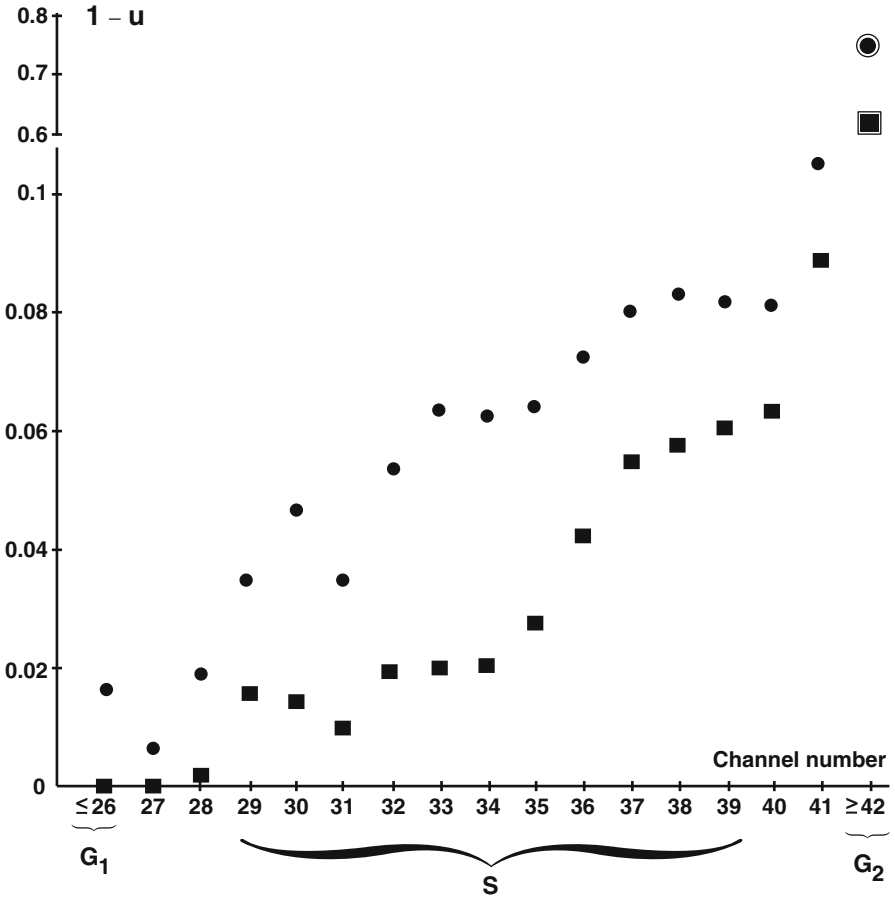


Fig. 6.16 Drug action curves for the low (10 nM, squares) and high (50 nM, circles) drug concentration. Fractions of cells blocked by the drug in given cell cycle compartments are plotted against corresponding numbers of the flow cytometer channels. The G_2 fractions (channel $\geq$ 42) are depicted on a different scale. (Source: Kimmel and Traganos1986)

in M, $T_M = 0.60$ h; and the growth rate (Malthusian parameter), $b = 0.062$ h^{-1}, corresponding to the doubling time of 11.22 h.

The fractions of cells blocked by the drug in the cell cycle compartments defined in the previous section were computed from the drug-perturbed stathmokinetic data. They are presented in Fig. 6.16, for low (10 nM) and high (50 nM) concentration of the drug. As evident from this graph, the blocking action of the drug is higher for cells more advanced in their progression through S. The durations (T_j) of the 15 successive subcompartments of the S phase are not very different from each other (mean duration:0.34 h; coefficient of variation: 0.13)

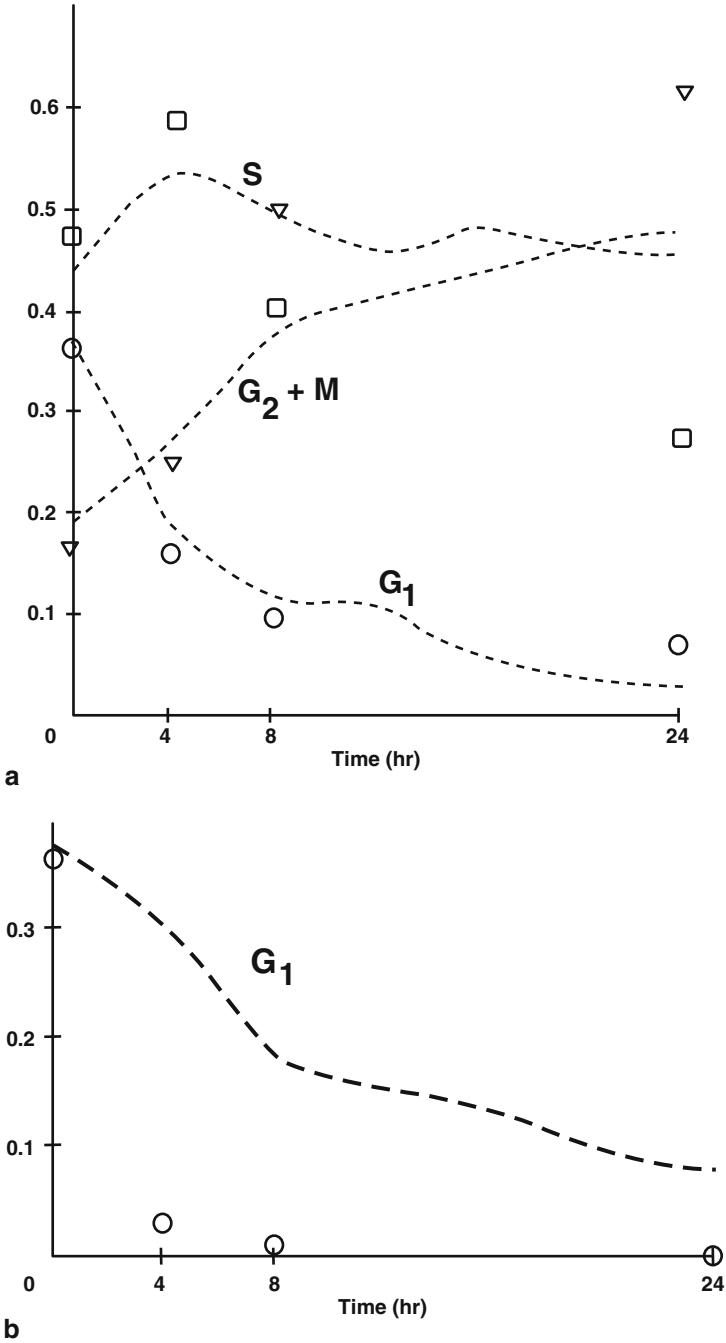


Fig. 6.17 Observed versus measured cell count fractions in different phases of the cell cycle, under continuous exposure to the drug: (**a**) low concentration (10 nM), (**b**) high concentration (50 nM). Measurements: circles, G_1; squares, S; triangles, $G_2 + M$. (Source: Kimmel and Traganos 1986)

Fits to the stathmokinetic data obtained using the low concentration drug action curve of Fig. 6.16 are presented in Fig. 6.13. For the high drug concentration, the quality of the fit is similar.

Modeling of cell kinetics under continuous exposure to the drug, employing the drug action curves estimated from the stathmokinetic experiment (Fig. 6.16), is depicted in Fig. 6.17. For the low drug concentration the model of Fig. 6.15 provides an excellent prediction of the observed G_1 cell count fraction, Fig. 6.17a. The S and $G_2 + M$ cell count fractions are not so well modeled, though the general trend is reproduced. Modeling of the high drug concentration effects is not as successful, Fig. 6.17b.

6.4.5 Discussion

In theory, it should be possible to improve chemotherapeutic treatment of cancer by appropriately scheduling the administration of cytotoxic agents. An optimum schedule could, for example, take advantage of differences in cell cycle length of tumor and "critical" (sensitive) normal tissues, to affect the malignant cells concentrated in a different part of the cell cycle. However, interest in such proposals has diminished. As early as two decades back, Tannock (1978) has commented that: "Enthusiasm for this approach has varied from euphoria to despair."

It appears however that the problem might be reconsidered. Theoretical calculations (Dibrov et al. 1983, 1985, and references therein) indicate that the potential for improvement in treatment outcome due to chemotherapy scheduling may be considerable. One of the difficulties is in obtaining estimates of numerical parameters characterizing cell kinetics under the action of cytotoxic agents. It seems probable that abandoning the efforts to find the optimum scheduling of chemotherapy was caused largely by the inability to find good estimates of the parameters mentioned above.

The failure to predict effects of the long-term (continuous) exposure to the drug at the higher concentration (see Fig. 6.16b) is probably related to considerable cell damage at this concentration. This damage is not apparent in the course of the stathmokinetic experiment (in fact the drug action curves for the two drug concentrations differ only slightly), but it probably manifests itself during subsequent cell divisions.

6.5 Application: Cell Surface Aggregation Phenomena

This model is taken from the book by Macken and Perelson (1985). Molecules on the cell surface (receptors) are activated by contact with molecules located in the extracellular medium (ligands). The activated receptors initiate signalling pathways within cells resulting in cell proliferation and cell differentiation. The strength of the signals depends on the specificity of the interaction between a ligand and its receptor, and the number of activated receptors per cell.

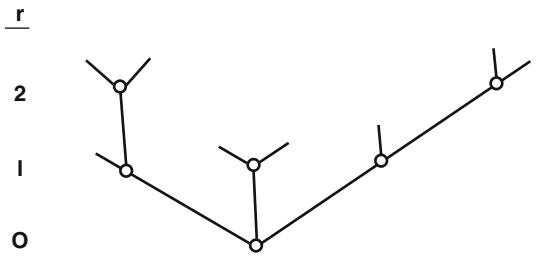


Fig. 6.18 A typical family tree representing the aggregation of f-valent particles. Here $f = 3$. Notice particles in generation $n = 0$ can have at most f offspring, whereas in all later generations a particle can have at most $f - 1$ offspring. (Source: Macken and Perelson 1988)

Examples of ligands are hormones such as insulin and growth factors, and antigens such as proteins on the surface of bacteria and viruses. Some receptors and some ligands are multivalent, e.g., the receptor molecules can react with more than one ligand at a time and the ligands can react with more than one receptor at a time. Multivalency may result in clusters of ligand–receptor complexes. It is of interest to determine the size distribution of the aggregates, and the probability that they will continue to increase in size or stop increasing in size.

6.5.1 Relationship Between the Galton–Watson Process and the Aggregation Process

Let us suppose for the beginning that we are given a collection of m-valent particles of single type (an m-valent particle is one that can bind m other particles). We restrict our attention to the aggregates of these particles that contain no loops and hence have the topological form of a tree. We equate the probability of k particles being bound to a given particle, with p_k, the probability of this particle having k offspring. The particle valency in the aggregation process is accounted for in the Galton–Watson process by imposing a restriction on the maximum possible number of offspring contributed by a single parent to the next generation. Thus, a parent in generation 0 can have at most m offspring, whereas a parent in later generations can have at most $m - 1$ offspring, because one particle site is used to attach the particle to its own parent (Fig. 6.18).

To summarize, the analogy between Galton–Watson process and aggregation processes is that an n-mer is represented by a rooted tree containing n nodes, with the degree of the root being at most m and the degree of all other nodes being at most $m - 1$.

The purpose of the mathematical representation is to find the distribution of the sizes of aggregates. The total size Y of the aggregate is equal to the summary number

of objects produced in all generations of the Galton–Watson process:

$$Y = \sum_{n=0}^{\infty} Z_n. \tag{6.61}$$

We are interested in the distribution of random variable Y including cases when Y is infinite. This last possibility corresponds to the so called *gelation* in which the aggregation process escapes control and utilizes all the particles suspended in the medium. Let us note that Y can be finite only if the process $\{Z_n, n = 1, 2, \dots\}$ dies out with probability one, that is, in the subcritical and critical case. In the supercritical process, there exists the nonzero probability $1 - q$ that the number of generations is infinite. This latter is the probability of gelation.

6.5.2 Progeny Distributions

We have to specify p_k, the probability that k sites of a randomly chosen m-valent particle are bound. Let p be the probability that a randomly chosen site is bound. Then, because sites act independently, p_k is given by the binomial formula. Consequently, the progeny pgf in the zeroth generation is

$$f_0(s) = (ps + 1 - p)^m,$$

while in the succeeding generations, it is

$$f(s) = (ps + 1 - p)^{m-1}.$$

6.5.3 Antigen Size Distribution on a Cell Surface

We consider a model for multivalent antigens binding to and cross-linking bivalent cell surface receptors, following Macken and Perelson (1985). The model describes production of antibody by antigen-stimulated B lymphocytes.

Antigen particles (Fig. 6.19), present in the solution surrounding a population of cells, can bind at any of $m_\mathrm{a} = 3$ (out of 6 existing) binding sites to one free site of a cell surface receptor. Receptors are bivalent, they have two binding sites, i.e., $m_\mathrm{r} - 1 = 1$. The antigen, once bound to a receptor, may bind another single receptor site at any out of remaining $m_\mathrm{a} - 1 = 2$ sites or it may bind two free sites of *two* receptors. In the model, it is not allowed for the two antigen sites to bind to two sites of a *single* receptor, since this would violate the tree structure. Repeated binding creates patches of antigen particles cross-linking receptors on the cell surface. Gelation is equivalent to the formation of "infinite-size" (very large) antibody–receptor clusters on cell surface.

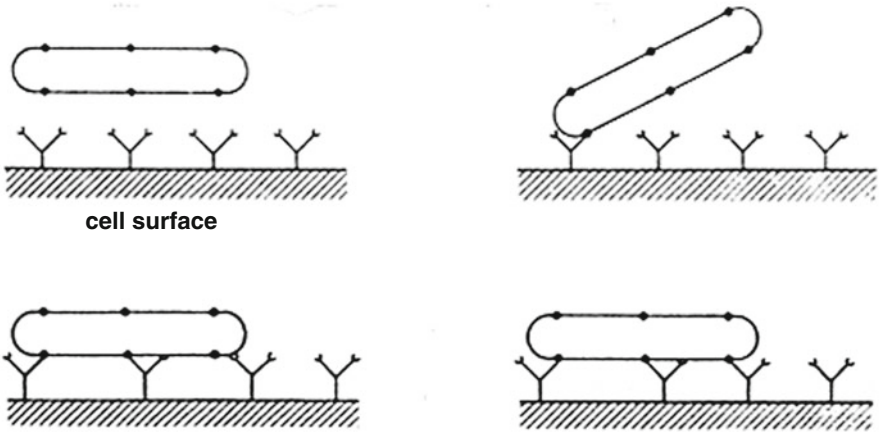


Fig. 6.19 A model for multivalent antigens binding to and cross-linking bivalent cell surface receptors. The antigen present in the solution surrounding a population of cells, can bind at any of $f = 3$ (out of 6 existing) binding sites to one site of a free cell surface receptor. Receptors are bivalent, i.e., have two binding sites. The antigen, once bound to a receptor, may bind another receptor at any out of remaining $f - 1 = 2$ sites, etc. (Source: Macken and Perelson 1988)

The special type of aggregate described above is distinguished by the fact that antigens and antibodies alternate along any path through the aggregate. Consequently, the model is described by a two-type Galton–Watson process $\mathbf{Z}_n = (Z_n^1, Z_n^2)$, $n = 1, 2, \ldots$, in which the offspring of the type 1 object (receptor particle) is of type 2 only and the offspring of the type 2 object (antigen particle) is of type 1 only, i.e.,

$$f^1(\mathbf{s}) = f^1(s_2) = (p_1 s_2 + 1 - p_1), \tag{6.62}$$

$$f^2(\mathbf{s}) = f^2(s_1) = (p_2 s_1 + 1 - p_2)^2. \tag{6.63}$$

We suppose that the process (aggregate) is started by a single receptor particle and therefore, for the zeroth generation,

$$f_0^1(s_2) = (p_1 s_2 + 1 - p_1)^2. \tag{6.64}$$

Calculations based on (6.62)–(6.64) show that the pgf $F_n(\mathbf{s})$ of the vector (Y_n^1, Y_n^2) of the counts of all particles of both types, up to generation n,

$$(Y_n^1, Y_n^2) = \sum_{i=0}^{n} (Z_i^1, Z_i^2),$$

is equal to

$$F_n(\mathbf{s}) = s_1 f_0^1\{s_2 f^2[s_1 f^1(s_2 \cdots)]\}. \tag{6.65}$$

The consequence of Eq. (6.65) is that the pgf $F(\mathbf{s})$ of the vector $(Y^1, Y^2) = \lim_{n\to\infty}(Y_n^1, Y_n^2)$ of the aggregate totals of particles of both types, is equal to

$$F(\mathbf{s}) = s_1 f_0^1[\Phi(\mathbf{s})],$$

where $\Phi(\mathbf{s})$ is the solution of the equation

$$\Phi(\mathbf{s}) = s_2 f^2\{s_1 f^1[\Phi(\mathbf{s})]\}. \tag{6.66}$$

The pgf solution of Eq. (6.66) always exists because of the monotone convergence. It may correspond to infinite particle count, i.e., gelation, if $\Phi(1,1) < 1$.

Obtaining an explicit expression for $\Phi(\mathbf{s})$ is possible. It is left as an exercise. We will derive the condition of supercriticality, and the probability of gelation for the supercritical process. The expected progeny matrix $\begin{pmatrix} 0 & p_1 \\ 2p_2 & 0 \end{pmatrix}$ is not positively regular (it has period 2) but it has a dominating root $\rho = (2p_1p_2)^{(1/2)}$. Thus, the criticality parameter is proportional to the geometric mean of the reactivities p_1 and p_2. The probability of gelation is equal to $1 - (1 - p_1)^2 q_2$, where $q_2 = (1 - p_1p_2)^2/(p_1p_2)^2$ is obtained by solving equation $(q_1, q_2) = [f^1(q_1, q_2), f^2(q_1, q_2)]$.

The above expressions are valid in the supercritical case.

6.6 Sampling Formulae for Multitype Galton–Watson Process

The literature on multitype branching processes is mostly focused on the asymptotic theory. In comparison, relatively little has been done to address problems of sampling in finite time from a branching process. This is a problem which is relevant in many biological applications. In the polymerase chain reaction (PCR), genetic material is amplified and sampled after a fixed number of cycles. In cell cultures, cells are grown and harvested after a fixed number of population doublings. Also, many branching processes arising in these applications are intrinsically reducible in the sense that some types can only have certain other types in their ancestries. In such processes, limiting distributions on the type space are typically degenerate and of no practical use.

In this section, we will present results by Olofsson and Shaw (2002) concerning sampling distributions in the multitype Galton–Watson process. These results allow us to find the expectation and variance of the frequency of particles of a given type in generation n of the multitype Galton–Watson process. These are given in terms of the probability generating function of the offspring distribution. Furthermore, given a particle of some type is sampled in generation n, the sequence of types of its parent particles in generations $n - 1$, $n - 2$, …, 2, 1, 0, is a discrete inhomogeneous Markov chain with different transition probabilities at each step. These results simplify simulations of genealogies and accumulated mutations in at least two interesting biological models (Sects. 6.7 and 6.8).

The approach taken in Olofsson and Shaw (2002) is similar to that used in a sequence of papers by Waugh (1981) and Joffe and Waugh (1982, 1986a, b), who address the so-called kin number problem in Galton–Watson populations. They establish exact formulas for the probability distributions of family trees of a randomly sampled individual in a fixed generation. The most extensive treatment is of the single-type case, Joffe and Waugh (1982), the multitype case is addressed in Joffe and Waugh (1986a, b).

We will use the notation ψ_n for the probability generating function of $(Z_n^0, \ldots, Z_n^r)$ when there is an arbitrary number of ancestors $(Z_0^0, \ldots, Z_0^r)$, reserving the notation f_n^i for the case of one single ancestor of type i.

6.6.1 Formulae for Mean and Variance

The following result gives the mean and variance of the proportion of type i individuals in the nth generation, conditioned on this generation being nonempty: We use the notation $|\mathbf{Z}_n|$ for the total number of individuals in the nth generation, i.e., $|\mathbf{Z}_n| = \sum_{k=0}^r Z_n^k$.

Theorem 6.5 *Let* $\mathbf{u}$ *be a vector with all* u *entries except for a* v *in the* ith *position:* $\mathbf{u} = (u, \ldots v, \ldots, u)$, $\mathbf{0} = (0, 0, \ldots, 0)$ *and denote by* ψ_n *the joint probability generating function of* $(Z_n^1, \ldots, Z_n^r)$. *Then*

$$\mathrm{E}\left[\left.\frac{Z_n^i}{|\mathbf{Z}_n|}\right| |\mathbf{Z}_n| > 0\right] = \frac{1}{1-\psi_n(\mathbf{0})}\int_0^1 \left.\frac{\partial}{\partial v}\psi_n(\mathbf{u})\right|_{u=v=s} ds$$

and

$$\mathrm{Var}\left[\left.\frac{Z_n^i}{|\mathbf{Z_n}|}\right| |\mathbf{Z}|\right] = \frac{1}{1-\psi_n(\mathbf{0})}\int_0^1 -\log s \left(s\left.\frac{\partial^2}{\partial v^2}\psi_n(\mathbf{u})\right|_{u=v=s} + \left.\frac{\partial}{\partial v}\psi_n(\mathbf{u})\right|_{u=v=s}\right) ds - \left(\frac{1}{1-\psi_n(\mathbf{0})}\int_0^1 \left.\frac{\partial}{\partial v}\psi_n(\mathbf{u})\right|_{u=v=s} ds\right)^2.$$

The methods of proof are inspired by those of Joffe and Waugh (1986) and Waugh (1981). Details of the proof are described in Olofsson and Shaw (2002).

6.6.2 The Markov Property

Next, we investigate the dependence structure in the sequence of types in the lineage of a particle in the nth generation. We may think of this particle as sampled at random

and denote its type by T_n. Since

$$\mathrm{P}(T_n = i) = \mathrm{E}\left[\frac{Z_n^i}{|\mathbf{Z}_n|} \middle| |\mathbf{Z}_n| > 0 \right]$$

the probability $\mathrm{P}(T_n = i)$ can be obtained from Theorem 6.5. Denote the type of this particle's parent by T_{n-1}, its grandparent's type by T_{n-2} and so on; we thus obtain a sequence of types $T_n, T_{n-1}, \ldots, T_0$, the type of the ancestor. It turns out that, conditional on nonextinction, this sequence is a nonhomogeneous Markov chain with transition probabilities given by a formulas invoking the probability generating function of the offspring distribution. This can be utilized for simulations to assess the type variation in lineages of sampled particles. Let φ_{ij} denote the probability generating function of the number of j-type offspring of an i-type parent, i.e.,

$$\varphi_{ij}(s) = \mathrm{E}_i[s^{X^{(j)}}] = f^i(1, \underbrace{\ldots, s, \ldots}_{j\text{th argument}}, 1)$$

and let ψ_k be as in Theorem 6.5.

Theorem 6.6 *The sequence of types $T_n, \ldots, T_0$ in the genealogy of an individual randomly sampled from generation n is a nonhomogeneous Markov chain with transition probabilities*

$$P(T_k = i | T_{k+1} = j) = \frac{1}{1 - P\left(Z_{k+1}^{(j)} = 0\right)} \int_0^1 \frac{\partial}{\partial v} \psi_k \left(\varphi_{0j}(u), \ldots, \varphi_{ij}(v), \ldots, \varphi_{rj}(u)\right) \Big|_{u=v=s} ds$$

where

$$P\left(Z_{k+1}^{(j)} = 0\right) = \psi_k \left(\varphi_{0_j}(0), \varphi_{1_j}(0), \ldots, \varphi_{r_j}(0)\right).$$

*Note that there is a v in the i*th *position and u in the other positions in the argument of ψ_k.*

Details of the proofs of both theorems are described in Olofsson and Shaw (2002). If the branching process is irreducible, then the backwards Markov chain becomes asymptotically homogeneous in the sense that as $n \to \infty$, the transition probabilities converge to limiting distributions. This follows from the convergence theorem of Jagers (1992) where convergence towards the so-called stable population is investigated. In the PCR application in Sect. 6.8, this can be observed empirically already for low values of n.

6.7 Application: Deletions in Mitochondrial DNA

Mitochondria are organelles in cells carrying their own DNA. Just like nuclear DNA, mitochondrial DNA (mtDNA for short) is subject to mutations which may take the form of base substitutions, duplications, or deletions. This application focuses on

one particular mutation, the mtDNA^{4977} deletion. This is a mutation which causes a deletion of about one third of the mitochondrial genome, thus causing a DNA molecule which is significantly smaller than normal. It has been observed that high levels of deletions are associated with certain degenerative diseases, for example Kearns–Sayre syndrome (Chinnery and Turnbull 1999). These levels may be as high as 40–50 %. Low levels (0.5–12 %) have been observed in different regions of the brain of healthy humans. There is a wide variety of issues involved such as different levels in different types of tissue but we will not attempt to address any of these. Instead, we focus on how the process of replication of mDNA can be described as a multitype Galton–Watson process and how the sampling formulas of the previous sections can be applied to explore how deletions accumulate over time. The idea to use branching processes in this application was first described in the unpublished manuscript of Navidi et al. (2003).

The population of mDNA is modeled as a two-type process where the types are 0 (normal) and 1 (mutant). A normal can give birth to either two normals or, if there is a mutation, one normal and one mutant. The latter happens with probability λ and we refer to λ as the mutation rate. Mutants can only give birth to mutants. A DNA molecule also may die without reproducing (so-called mitochondrial turnover, see Arking 1998) and we let the survival probabilities be p and q for normals and mutants, respectively. This gives the following offspring distributions:

$$p_0(0,0) = 1 - p, \quad p_0(2,0) = p(1-\lambda), \quad p_0(1,1) = p\lambda$$

for normals and

$$p_1(0,0) = 1 - q, \quad p_1(0,2) = q$$

for mutants. This gives the joint probability generating functions

$$\varphi_0(u,v) = 1 - p + p\lambda uv + p(1-\lambda)u^2 \tag{6.67}$$

and

$$\varphi_1(u,v) = 1 - q + qv^2. \tag{6.68}$$

The proportion of mutants in the nth generation is

$$\frac{Z_n^{(0)}}{Z_n^{(0)} + Z_n^{(1)}}$$

and we can use Theorem 6.5 to compute its expectation and variance. Further details are described in Olofsson and Shaw (2002).

6.8 Application: Polymerase Chain Reaction

This application can be understood as a sequel to Sect. 1.2. As described in that section, a DNA molecule in any given cycle of PCR either existed before that cycle or is newly created (this is the essence of the semiconservative replication). The process is modeled as a two-type process where the type space is $\{0, 1\}$, "0" for "old" and "1" for "new". The distinction is crucial to mutation studies since new mutations only arise on newly created particles. The offspring distribution is

$$p_0(1,0) = 1 - p, \quad p_0(1,1) = p,$$
$$p_1(1,0) = 1 - p, \quad p_1(1,1) = p,$$

where p is the cycle efficiency, i.e., the probability that a given molecule replicates successfully in a given PCR cycle. This leads to joint probability generating functions

$$\varphi_0(u,v) = \varphi_1(u,v) = (1-p)u + puv.$$

For the simulations, Theorem 6.5 was used to compute the distribution of a randomly sampled particle in generation n, and Theorem 6.6 to compute the transition probabilities. Simulations were then performed in which a particle was sampled at random from generation n and the sequence of types in its lineage back to the ancestor was generated. Each time a particle of type 1 appeared, it was independently assigned a new mutation with probability λ. The values $n = 30$, $p = 0.7$, and $\lambda = 0.05$ were used (see Sect. 1.2, where however a slightly different notation, consistent with the Weiss and von Haeseler (1997) paper, was used).

Olofsson and Shaw (2002) show a histogram of the number of mutations in the lineage of a randomly sampled particle in generation 30, based on 100,000 simulation runs of the Markov chain. The transition probabilities $P(T_k = i | T_{k+1} = j)$ converge to a limiting distribution as $n \to \infty$ and in this particular application, the convergence is rapid. The limiting transition probabilities can be computed as

$$P(T_k = i | T_{k+1} = j) = \frac{\nu(i) M(i,j)}{\rho \nu(j)}$$

where $M(i,j) = E_i[X^{(j)}]$, the (i,j)th entry in the mean reproduction matrix

$$M = \begin{pmatrix} M(0,0) & M(0,1) \\ M(1,0) & M(1,1) \end{pmatrix} = \begin{pmatrix} 1 & p \\ 1 & p \end{pmatrix},$$

ρ is the largest eigenvalue of M and ν is the left eigenvector of M corresponding to ρ. In this case,

$$\rho = 1 + p, \quad \nu(0) = \frac{p}{1+p}, \quad \nu(1) = \frac{1}{1+p},$$

which gives, in the limit

$$P(T_k = 1|T_{k+1} = j) = \frac{p}{1+p} \approx 0.41,$$

for both $j = 0$ and 1. The computations reveal that this limit is reached after less than ten generations. Still further details may be found in the unpublished dissertation by Shaw (2000).

6.9 Other Works and Applications

6.9.1 Hemopoiesis and Clonal Cell Populations

Multitype branching processes are the natural tool to model proliferation of cells undergoing differentiation, i.e., changing gene expression, morphology, and biological function. Usually, as it is the case in the hemopoietic (blood production) system, populations of differentiating cells are organized in nets. The cells at the top of the net are stem cells. They can produce progeny of their own type or of an other type. Each next population is committed to differentiation in some direction, i.e., it can produce progeny of its own type or of a limited subset of types, usually just one type of more mature cells. The bottom population(s) are not capable of proliferation. This type of multitype branching process is called reducible. Early papers employing branching-type models are Till et al. (1964) and Vogel et al. (1969). A simulation model was developed by Rittgen (1983).

The stochastic model of mast cell proliferation developed by Pharr et al. (1985) assumed a two-type Galton–Watson process including proliferative and nonproliferative cells. Each proliferative cell gives rise to either two proliferative progeny, or two nonproliferative progeny, or it may die. Each nonproliferative cell may either survive (i.e., continue to exist as 1 cell) or die. In principle, this model is identical to that of Sect. 3.2. Predictions of the model by Pharr et al. (1985) were fitted to colony size data, with a good agreement. In a further paper (Nedelman et al. 1987), the model was extended to a Bellman–Harris process and maximum likelihood was used to estimate parameters.

A more general model including a chain of maturing cell populations, of the type described above, was designed by Ciampi et al. (1986) to model proliferation of human ovarian carcinoma cells. Modeling, using a multitype Galton–Watson process involved calculating the asymptotic distributions of colony sizes and data-based estimation of the self-renewal probability of stem cells. This latter is the conditional probability of a stem cell producing two stem cells (as opposed to producing two differentiated cells) given it does not die or rest. The self-renewal probability is a parameter of potential diagnostic value. Another related reference is the book by Macken and Perelson (1988), which considers multitype Galton–Watson models of the hemopoietic (blood production) system although without much reference to

data. Therneau et al. (1989) model early stages of development of cell colonies using symbolic calculations to iterate the probability generating functions of the process.

Papers by Stivers and Kimmel (1996a, b) and by Stivers et al. (1996) concerned the observed inheritance of sizes of primary and secondary colonies in experiments by Axelrod et al. (1993) and Gusev and Axelrod (1995), discussed in Sect. 5.5.1. The main question is to determine what modes of inheritance of cell lifetimes are consistent with the observed correlations of the sizes of primary and secondary colonies, which are positive, equal to approximately 0.6 and, at the same time, consistent with the observed variances in colony sizes. Various modes are considered, including "clonal," in which the lifelength distribution of the secondary colony is affected by the lifetime of its founder drawn at random from a primary colony (Stivers and Kimmel 1996a, b). Another variant is generational inheritance in a model where two types of cells, with differing proliferative potential, can differentiate into each other (Stivers et al. 1996). This latter model provides a fit to the observations.

Abkowitz et al. (1996) use experimental data and branching process simulations to demonstrate that hemopoiesis, the process of blood cell production, has a random nature. They use results of irradiation experiments carried out on Safari cats, a race being a cross between domestic cats and wild Geoffroy cats. These two species of cats have evolved independently and have electrophoretically distinct phenotypes of the X-chromosome-linked enzyme glucose-6-phosphate dehydrogenase (G6PDH). Female Safari cats are generally balanced heterozygotes with, on average, equal numbers of progenitor and differentiated blood cells of each parental phenotype. However, females deprived of their bone marrow by irradiation and then given autologous transplants of 30 quiescent hematopoietic cells, end up, after a period of fluctuations, with variable proportions of progenitors of each parental phenotype. The pattern of variability is consistent with simulations based on a multitype branching process model. This paper, although it contains no mathematics, provides important arguments concerning applicability of branching processes.

6.9.2 Gene Amplification

The biological introduction to gene amplification can be found in Sect. 2.1.6. Mathematical models are described in Sects. 3.7, 6.1, 5.4, and 7.4. Other authors, considered diverse aspects of gene amplification. Harnevo and Agur (1991) constructed a comprehensive model of gene amplification in the form of a multitype branching process, in the context of resistance of cancer cells to cytotoxic drugs. The number of types is finite, as these authors assume limits to the number of copies of the amplified gene. Theoretical considerations are followed by a modeling study in which the dynamics of growth of cells with amplified phenotype (drug resistant) is followed. A similar mathematical model was employed by Harnevo and Agur (1992) to explore various strategies of cancer chemotherapy, assuming that the main mechanism of drug resistance was gene amplification. The paper by Harnevo and Agur (1993) contains

a critical discussion of approaches to modeling of the gene amplification process, including the model of Kimmel et al. (1992), described in Sect. 7.5.

Finally, we should mention the paper by Peterson (1984). This is a paper in which evidence is collected suggesting that expression of many proteins in cells occurs at levels which form an arithmetic or geometric progression. Peterson (1984) postulates that this may be due to variable number of copies of respective genes, present in a given cell (quantitative shift model). Although this paper does not contain mathematics, a multitype branching process is implicitly involved.

6.9.3 Modeling in Varying Environments

Modeling of multitype branching processes in varying environments is important when we consider populations of biological cells subject to external controls. The usual backward (retrospective) technique of decomposition into subprocesses generated by first-generation progeny leads to pgf and moment equations which tie together processes started at different times and therefore different, so that self-recurrence cannot be used. This is inconvenient if the solution is to be extended step-by-step, for example by numerical procedures, as it is the case in modeling of cancer chemotherapy. For this reason, it is desirable to develop a forward (prospective) technique, which would provide a recurrence or equation allowing to continue in time the pgf or moment characterization of the process. In Kimmel (1982), it is demonstrated that an equivalent (dual) set of integral equations exists, which allows prospective continuation of the expectations of the process. In Kimmel (1983), it is shown that a prospective equation of a kind can be written not for the probability generating functions but for the probability generating functional which describe the multivariate point process of births and deaths in the branching process.

6.9.4 Model of Ovarian Cancer Progression and Metastasis

Danesh et al. (2012) have devised a branching process model of growth and progression of ovarian cancer, which is one of the major killers among hormone-dependent cancers. The reason it is so deadly is that it is detected relatively late; only 32% detected cases are metastasis-free at diagnosis. The model is developed in the terms of a nonstandard multitype branching process, which describes progression of the tumors over different primary and metastatic stages. The paper contributes to public-health literature in that it allows computing rigorously the "window of opportunity" for screening, during which the cancer is detectable but still early enough to be treated for cure. The width of this window is about 2.9 years, so patients should be screened at about 2-year intervals for early detection to be accomplished.

6.9.5 HIV Modeling

Conway and Coombs (2011) use a Markov age-dependent branching process with three types of particles, to model probability distributions of the number of virus particles in HIV infection treated using antiretroviral compounds. The motivation of the study is viral persistence in HIV+ patients on long-term antiretroviral treatment (ART). The stochastic model of HIV viral dynamics in the blood stream is based on the hypothesis that the residual viremia in patients on ART can be explained by the activation of cells latently infected by HIV before the initiation of ART and that viral blips (clinically observed short periods of detectable viral load) represent large deviations from the mean. Modeling such deviations requires a stochastic approach. Blip amplitudes and frequencies are calculated by computing complete viral load probability distributions, and the duration of viral blips is studied via direct numerical simulation. The model qualitatively reproduces short small-amplitude blips detected in clinical studies of treated HIV infection.

In an unrelated study, Shiri and Welte (2011) use a branching process model in a deterministically varying environment to explore how dynamics of early HIV infection impact on accumulation of mutations driving the long-term evolution of drug resistance and immune system evasion. The branching process theory offers the ability to compute important indicators of viral diversity, in a framework with a limited number of simplifying assumptions, without the need to simulate the full range of individual level scenarios. These models may be useful to illustrate the impact of vaccines and treatments on viral evolution. They also suggest that new measures of viral diversity which correlate to prognosis should be sought in therapy and vaccination trials.

Chapter 7
Branching Processes with Infinitely Many Types

In this chapter, we consider a number of examples of branching processes with infinite type spaces. No systematic theory can be presented. However, in Sect. 7.1 we review various approaches generalizing the denumerable case. Also, general processes (Sect. C.1) include the denumerable type space as a special case. We will base considerations on an analogy with the finite mutitype case whenever possible. However, the stress is on interesting and diverse properties, which are different from the finite multitype setup and on biologically motivated examples.

We will begin by presenting an example of a stable process, using another variant of the gene amplification model (Sect. 7.5). The subsequent example is the reducible process of loss of telomere sequences at the end of chromosomes, which displays a polynomial dynamics (Sect. 7.7). Sections 7.4–7.6 deal with the problem of quasistationarity in the context of branching random walks and branching-within-branching. These examples can be understood as generalizations of the Yaglom Theorem 3.6, in this book. Finally, in Sect. 7.8, we develop a series of structured population models, which can be classified as branching processes with continuous type spaces.

7.1 Galton–Watson and Bellman–Harris Processes with Denumerably Many Types and Branching Random Walks

One of the questions arising when a multitype process is generalized to denumerable infinity of types is, which of the simple properties of the finite case remain valid. The answer is of importance for the applications, since it helps to decide what new properties are expected from a model, if the constraint on the number of types is released. Applications in several sections in the present chapter demonstrate that in general a variety of asymptotic behaviors can be expected.

One of the papers devoted to this issue is Spătaru (1989). This paper considers the extension of the multitype Galton–Watson (GW) process to countably many types indexed by $\mathbf{N}$. Let $Z_n = (Z_{n\alpha})$ be the vector of generation sizes; $Z_{n\alpha}$ is the number of α-types in generation n ($\alpha \in \mathbf{N}$). Let $M = [M_{\alpha\beta}]$, where $M_{\alpha\beta} = \mathrm{E}(Z_{1\beta}|Z_{0\beta} = \delta_{\alpha\beta}$, $\beta \in \mathbf{N}$), be the mean matrix, and let $f(s) = (f_\alpha(s))$, where, for $s \in C = [0,1]^{\mathbf{N}}$,

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology*,
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_7

$f_\alpha(s) = \mathrm{E}(s^{Z_n} | Z_{0\beta} = \delta_{\alpha\beta},\ \beta \in \mathbf{N})$. The author shows that the nonzero states are transient if M is irreducible and $f(s) \neq Ms$ for some $s \in C$. It is asserted that transience does not imply the property $\mathrm{P}(Z_n \to 0 \text{ or } |Z_n| \to \infty) = 1$, valid when the number of types is finite.

Let q denote the vector of extinction probabilities. If $f_n(s)$ denotes the vector-valued n-fold functional iterate of $f(s)$ (i.e., $f_{\alpha n}(s) = f_\alpha(f_{n-1}(s))$), then $q = \lim_{n\to\infty} f_n(0)$. For which s, is it true that $f_n(s) \to q$? He shows it is true if $s \leq q$ (coordinatewise), but not true for all $s \in C\backslash\{\mathbf{1}\}$, where $\mathbf{1} = (1, 1, \cdots)$, if $S = \{s \in C : s = f(s)\}$ has number of elements exceeding 2. If M is irreducible and f is not affine, then the number of elements of S is equal to 2 and $q = \mathbf{1}$ or $q < \mathbf{1}$

Another paper on a related subject is Moy (1967). A denumerable-type GW process is considered, with mean progeny matrix $M = (m_{ij})$, where $m_{ij} = \mathrm{E}[Z_{n+1}(j) | Z_n = e_i]$. The principal role is played by the Perron–Frobenius root r of M, in this case, the radius of convergence of the power series $\sum_i M^i s^i$. The Perron–Frobenius root plays the role of the reciprocal of the Malthusian parameter. Two cases are possible: (I) $\sum_i M^i r^i$ finite, and (II) $\sum_i M^i r^i$ infinite. In case I, there exist two strictly positive infinite sequences v and u, unique up to multiplicative constants, satisfying $rvM = v$ and $rMu = u$, i.e., left and right eigenvectors corresponding to eigenvalue r^{-1}. In case II, under an additional condition $\sum_i u(i)v(i) < \infty$, and if the process is supercritical, i.e., if $r^{-1} > 1$, $Z_n r^n$ converges in mean square to vY, where Y is a scalar random variable. In the remaining cases, $Z_n r^n$ converges to 0.

It seems that the conditions for asymptotic behavior of the supercritical process can be obtained as conclusions from the conditions for the general branching process of Sect. C.1. In the case of branching process with denumerable type space, these conditions seem to be a nontrivial extension of the positive regularity conditions sufficient in the finite multitype case (see Theorem 6.2). Indeed, the branching random walk of Sect. 7.4, conditional on nonentering the 0-state, is a supercritical branching process. Its expected progeny matrix is irreducible in the sense of two arbitrary states communicating in a finite number of steps. However, as seen from Theorem 7.2, the asymptotics conditional on nonentering the 0-state is exponential modified by a negative power multiplier, and not a pure exponential. This means that the reproductive kernel of this process cannot be conservative, in the sense of condition (C.4), although a direct proof seems nontrivial.

Kesten (1989) proves a limit theorem for the rate of growth of a supercritical multitype branching process with countably many types. He proves, under appropriate conditions, that both the growth rate and the direction of growth in type space are essentially deterministic. The principal motivation for this work is to extend branching process theory to a problem arising in the study of random fractals, i.e., the properties of the projections of random Cantor sets in d dimensions onto subspaces of small dimensions.

A large number of papers were written on the subject of branching random walks, i.e., denumerable-type branching processes with type-space transitions having the form of random walk. The typical problems considered include the rate of spread and growth of the branching random walk (Biggins 1995, Biggins et al. 1997) and

the Seneta–Heyde norming (Biggins and Kyprianou 1996). A surprisingly small number of papers are devoted to branching random walks with restrictions, of the type considered in Sect. 7.4, and other sections. One example is Biggins et al. 1991, considering a supercritical branching random walk on the real line commencing with a single ancestor at the origin. All individuals reproduce according to the same law with mean family size $b > 1$. Each progeny is given an independent identically distributed (iid) displacement from its parent with distribution F having negative mean and an exponentially decaying right tail, i.e., $\int_{-\infty}^{\infty} \mathrm{e}^{st} dF(t) < \infty$ for some s. The process is then attenuated by deleting all individuals below $(-x)$ and their descendants. Each remaining line of descent is just a random walk, starting at 0, with a barrier at $(-x)$, where $x > 0$. Results concerning the extinction probability and the expected population size depend on the parameter $h = \sup_\theta (-\log \int_{-\infty}^{\infty} \mathrm{e}^{\theta t} dF(t))$. Specifically, if $b < \mathrm{e}^h$, the probability that the process becomes extinct is 1 and, if $b > \mathrm{e}^h$, the probability of nonextinction is strictly positive. In the case $b < \mathrm{e}^h$ and F nonlattice, the expected size of the total population, denoted by $f(x)$, satisfies $\lim_{x\to\infty} \mathrm{e}^{-\alpha x} f(x) = C$, where α is the smallest positive solution of the equation $b \int_{-\infty}^{\infty} \mathrm{e}^{\alpha t} dF(t) = 1$ and C is a positive constant which can be estimated.

7.2 Generalized Linear-Fractional Distributions and Their Applications

7.2.1 Introduction

The current account of the generalized linear-fractional (LF) distributions is based on the papers by Sagitov (2011, 2013). The author studies multitype Bienaymé–Galton–Watson (BGW) processes with LF reproduction laws using various analytical tools such as contour process, spinal representation, and Perron–Frobenius theorem for countable matrices and renewal theory. For this special class of branching processes with countably many types he presents a transparent criterion for R-positive recurrence with respect to the type space. This criterion appeals to the Malthusian parameter and the mean age at childbearing of the associated LF Crump–Mode–Jagers process. Here, we will only summarize essentials that make the approach interesting.

A LF BGW process with countably many types is fully specified by a triplet of parameters $(\mathbf{H}, \mathbf{g}, m)$, where $\mathbf{H} = (h_{ij})_{i,j=1}^{\infty}$ is a substochastic matrix, $\mathbf{g} = (g_1, g_2, \dots)$ is a proper probability distribution, and m is a positive constant. For a given triplet $(\mathbf{H}, \mathbf{g}, m)$, the particles in the LF BGW process have the following reproduction law: A particle of type i has no offspring with probability $h_{i0} = 1 - \sum_{j\geq 1} h_{ij}$. Given that this particle has at least one offspring, the type of its first daughter is j, with probability $h_{ij}/(1 - h_{i0})$, and the number of subsequent daughters has a geometric distribution with mean m. With the exception of the first daughter the types of all other offspring particles follow the same distribution $\mathbf{g}$ independently of each other and *independently of mother's type*.

The countable matrix of the mean offspring numbers

$$\mathbf{M} = (m_{ij})_{i,j=1}^{\infty}, \quad m_{ij} = \mathbb{E}(Z_j^{(1)} | \mathbf{Z}^{(0)} = \mathbf{e}_i),$$

where $\mathbf{e}_i = (1_{\{i=1\}}, 1_{\{i=2\}}, \dots)$, in the LF case is found as

$$\mathbf{M} = \mathbf{H} + m\mathbf{H}\mathbf{1}^{\mathrm{t}}\mathbf{g}, \tag{7.1}$$

where $\mathbf{1}^{\mathrm{t}}$ is the transpose of the row vector $\mathbf{1}^{\mathrm{t}} = (1, 1, \dots)$.

Assume that the type of the initial particle has distribution $\mathbf{g}$. Then the total population sizes $Z^{(n)} = \mathbf{Z}^{(n)}\mathbf{1}^{\mathrm{t}}$ for the LF BGW process form a single-type discrete-time Crump–Mode–Jagers (CMJ) process (Jagers and Sagitov 2008), which is called a *LF CMJ process*. This CMJ model is not restrictive about the life length distribution, however, the point process of birth events must follow a very specific pattern: At each age a living individual produces independent and identically distributed (iid) geometric numbers of daughters. Making birth events in the LF CMJ process very rare (by choosing the row sums of $\mathbf{H}$ to be close to zero) and rescaling time accordingly we arrive at a continuous time CMJ process studied in Lambert (2010). In Lambert (2010), special attention is paid to the properties of the so-called contour processes of the corresponding planar genealogical trees.

7.2.2 Definitions and Basic Properties

The following vector notation is used: $\mathbf{x} = (x_1, x_2, \dots)$, $\mathbf{s}^{\mathbf{x}} = s_1^{x_1} s_2^{x_2} \dots$, $\mathbf{0} = (0, 0, \dots)$. Let $\mathbf{x}^{\mathrm{t}}$ stand for the transpose of the vector $\mathbf{x}$, and $\mathbf{I}$ denote the identity matrix $(1_{\{i=j\}})_{i\geq 1, j\geq 1}$. We denote by $\mathbb{Z}_+^{\infty}$, the set of vectors $\mathbf{k}$ with nonnegative integer-valued components and finite $k = \mathbf{k}\mathbf{1}^{\mathrm{t}}$.

Definition 7.1 *Let $(h_0, h_1, h_2, \dots)$ be a probability distribution on $\{0, 1, 2, \dots\}$, $(g_1, g_2, \dots)$ be a probability distribution on $\{1, 2, \dots\}$, and m be a positive constant. Put $\mathbf{h} = (h_1, h_2, \dots)$, $\mathbf{g} = (g_1, g_2, \dots)$. We say that a random vector $\mathbf{Z}$ has a LF distribution LF$(\mathbf{h}, \mathbf{g}, m)$ if*

$$\mathbb{P}(\mathbf{Z} = \mathbf{0}) = h_0, \quad \mathbb{P}(\mathbf{Z} = \mathbf{k} + \mathbf{e}_i) = \frac{h_i m^k}{(1+m)^{k+1}} \binom{k}{k_1, k_2, \dots} \mathbf{g}^{\mathbf{k}}$$

for all $\mathbf{k} \in \mathbb{Z}_+^{\infty}$, where $k = \mathbf{k}\mathbf{1}^{\mathrm{t}}$.

The name of the distribution is explained by the LF form of its multivariate generating function

$$\mathbb{E}(\mathbf{s}^{\mathbf{Z}}) = h_0 + \frac{\sum_{i=1}^{\infty} h_i s_i}{1 + m - m\sum_{i=1}^{\infty} g_i s_i}$$

which is an extension of its one-dimensional version.

Definition 7.2 *Let* $\mathbf{H} = (h_{ij})_{i,j=1}^{\infty}$ *be a substochastic matrix with rows* $\mathbf{h}_i = (h_{i1}, h_{i2}, \dots)$ *having nonnegative elements such that* $h_{i0} := 1 - h_{i1} - h_{i2} - \dots$ *take values in* $[0, 1]$. *Let* $\mathbf{g} = (g_1, g_2, \dots)$ *be a probability distribution on* $\{1, 2, \dots\}$, *and* m *be a positive constant. A multitype BGW process will be called LF with parameters* $(\mathbf{H}, \mathbf{g}, m)$, *if for all* $i = 1, 2, \dots$ *particles of type* i *reproduce according to the LF*$(\mathbf{h}_i, \mathbf{g}, m)$ *distribution.*

Simulation of $\mathbf{Z}$ using two dice with countably many sides each:

1. The $\mathbf{r}-$die gives the type of the first particle, if any
2. The $\mathbf{t}-$die generates the Geom (t_0) number of the remaining particles as well as their types

gives a representation in a slightly different form

$$E(\mathbf{s}^{\mathbf{Z}}) = r_0 + \frac{t_0 \sum_{i=1}^{\infty} r_i s_i}{1 - \sum_{i=1}^{\infty} t_i s_i}.$$

With this formula for a LF pgf as a starting point, a naive attempt to introduce a branching process with countably many types would define the reproduction law of type i particles by pgfs

$$E(\mathbf{s}^{\mathbf{Z}^{(1)}} | \mathbf{Z}^{(0)} = \mathbf{e}_i) = r_{i0} + \frac{t_{i0} \sum_{j=1}^{\infty} r_{ij} s_j}{1 - \sum_{j=1}^{\infty} t_{ij} s_j}$$

allowing for most general dependence on the mother's type. However, as it was shown in the finite-dimensional case (Pollak 1974) and (Joffe and Letac 2006) for the convolutions of the LF functions to be again LF, it is necessary that the parameters $t_j = t_{ij}$ are independent of the mother type i.

Theorem 7.1 *Consider a LF BGW process with parameters* $(\mathbf{H}, \mathbf{g}, m)$ *starting from a type* i *particle. Its* nth *generation size vector* $\mathbf{Z}^{(n)}$ *has a LF distribution* $LF(\mathbf{h}_i^{(n)}, \mathbf{g}^{(n)}, m^{(n)})$ *whose parameters satisfy*

$$m^{(n)} = m \sum_{k=0}^{n-1} \mathbf{g}\mathbf{M}^k \mathbf{1}^{\mathrm{t}}, \tag{7.2}$$

$$m^{(n)} \mathbf{g}^{(n)} = m\mathbf{g}(\mathbf{I} + \mathbf{M} + \dots + \mathbf{M}^{n-1}), \tag{7.3}$$

$$\mathbf{H}^{(n)} = \mathbf{M}^n - \frac{m^{(n)}}{1 + m^{(n)}} \mathbf{M}^n \mathbf{1}^{\mathrm{t}} \mathbf{g}^{(n)}, \tag{7.4}$$

where $\mathbf{H}^{(n)}$ *is the matrix with the rows* $(\mathbf{h}_i^{(n)})_{i=1}^{\infty}$.

7.2.3 Applications in Branching Processes

Multitype BGW processes are classified according to the asymptotic properties of the mean matrices $\mathbf{M}^{(n)} = (m_{ij}^{(n)})_{i,j=1}^{\infty}$ with elements

$$m_{ij}^{(n)} = \mathbb{E}(Z_j^{(n)} | \mathbf{Z}^{(0)} = \mathbf{e}_i)$$

as $n \to \infty$. The assumed independence of particles implies a recursion $\mathbf{M}^{(n)} = \mathbf{M}\mathbf{M}^{(n-1)}$, where $\mathbf{M} = \mathbf{M}^{(1)}$. It follows that $\mathbf{M}^{(n)} = \mathbf{M}^n$. Given that all powers $\mathbf{M}^n$ are element-wise finite (which is always true in the LF case) the asymptotic behavior of these powers is described by the Perron–Frobenius theory for countable matrices (see Chap. 6 in Seneta 2006).

Recall that a nonnegative matrix $\mathbf{M}$ is called irreducible, if for any pair of indices (i, j) there is a natural number n such that $m_{ij}^{(n)} > 0$. The period of an index i in an irreducible matrix $\mathbf{M}$ is defined as the greatest common divisor of all natural numbers n such that $m_{ii}^{(n)} > 0$. In the irreducible case, all such indices have the same period which is called the period of $\mathbf{M}$. When this period equals one, the matrix $\mathbf{M}$ is called aperiodic.

Due to Theorem 6.1 from Seneta (2006), all elements of the matrix power series $\mathbf{M}(s) = \sum_{n\geq 0} s^n \mathbf{M}^n$ have a common convergence radius $0 \leq R < \infty$, called the convergence parameter of the matrix $\mathbf{M}$. Furthermore, one of the two alternatives holds:

- R-transient case: $\sum_{n=0}^{\infty} m_{ii}^{(n)} R^n < \infty,\ i \geq 1$
- R-recurrent case: $\sum_{n=0}^{\infty} m_{ii}^{(n)} R^n = \infty,\ i \geq 1$

According to Seneta (2006; Theorem 6.2 and a remark afterwards), in the R-recurrent case, there exist unique up to constant multipliers *positive* vectors $\mathbf{u}$ and $\mathbf{v}$ such that

$$R\mathbf{M}\mathbf{u}^{\mathrm{t}} = \mathbf{u}^{\mathrm{t}},\ R\mathbf{v}\mathbf{M} = \mathbf{v}.$$

Using $Rv_j m_{ji}/v_i$, one can transform the matrix $\mathbf{M}$ into a stochastic matrix. The R-recurrent case is further divided into two subcases: R-null, when $\mathbf{v}\mathbf{u}^{\mathrm{t}} = \infty$, and R-positive, when $\mathbf{v}\mathbf{u}^{\mathrm{t}} < \infty$.

Definition 7.3 *A BGW process with countably many types will be called subcritical (critical, supercritical) and transient {recurrent, null-recurrent, positively recurrent} in the type space, if its matrix of the mean offspring numbers* $\mathbf{M}$ *has a convergence radius* $R > 1$ *(*$R = 1$, $R < 1$*) and is* R*-transient {*R*-recurrent,* R*-null recurrent,* R*-positively recurrent}*

Proposition 7.1 *In the supercritical positively recurrent case when the Perron–Frobenius eigenvalue of* $\mathbf{M}$, $\rho > 1$,

$$\mathbb{P}(\mathbf{Z}^{(n)} \neq \mathbf{0}) \to\ (\rho - 1)(1 + m)^{-1}\beta\mathbf{u}^{\mathrm{t}}.$$

Furthermore, for any $\mathbf{w}$ *with bounded components and* $\mathbf{v}\mathbf{w}^{\mathrm{t}} > 0$,

$$\mathbb{P}(\mathbf{Z}^{(n)}\mathbf{w}^{\mathrm{t}} > \rho^n x | \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_i) \to \mathrm{e}^{-x(\rho-1)/c_w}, \quad x > 0. \tag{7.5}$$

This result concludes our elementary review of Sagitov's theory. The paper by Sagitov (2013) contains a number of interesting results and proofs, some of them carried out using the methodology of branching process trees and contour processes. He also extends the results to the Crump–Mode–Jagers process. BGW processes with countable type space provide a natural mathematical tool for analysis of dynamics of proliferating particles within cells, such as viral genomes, amplified gene copies, and other. The process considered by Sagitov (2013) has not yet been applied to any biological model. However, despite the restrictive requirement of all progeny distributions being independent of parent type, the LF process is of interest for biological applications.

7.3 Biological Models with Denumerable Infinity of Types

An example of such an application is the paper by Taïb (1993), where a branching model is proposed for the behavior of populations of the budding yeast *Saccharomyces cerevisiae*. Using the idea of branching processes counted by random characteristics (Sect. C.1.2), explicit expressions are obtained describing different aspects of the asymptotic composition of such populations. Using the author's words, "The main purpose of this note is to show that the branching process approach is an alternative to deterministic population models based on differential equation methods."

A complementary reading to the material of Sect. 7.7 is a paper by Kowald (1997), which concerns the possible mechanisms for the regulation of telomere length. As mentioned in Sect. 7.7, since DNA polymerases can only synthesize a new DNA strand in the $5'$–$3'$ direction and needs a primer that provides a free $3'$ end, the cellular replication machinery is unable to duplicate the $3'$ ends of linear chromosomes unless special mechanisms are operative. While the telomeres seem to shorten continuously in human somatic cells because of the "end replication" problem, it appears that telomere length is maintained in cancer cells, the germ line, and unicellular organisms like yeast and *Tetrahymena* by a mechanism involving the enzyme telomerase, which elongates the $3'$ ends of telomeres. However, telomerase must be part of a more complicated mechanism to ensure that there is no net gain or loss of telomeric ends. Kowald (1997) describes a simple theoretical model, being in essence, a denumerable-type branching process that can explain several experimental findings. The simulations show that (i) the proposed mechanism is able to maintain telomeres at a constant length, (ii) this length constancy is independent of the initial telomere length, (iii) mutations of the telomeric sequence lead to an elongation of telomeres, (iv) inhibition of telomerase causes telomeric shortening, and (v) it reproduces and explains the experimental result that the addition of oligonucleotides to the culture medium leads to an increase of telomere length. Although no formal mathematical analysis is carried out by Kowald (1997), the model may lead to interesting applications.

7.4 Application: A Model of Unstable Gene Amplification

In this section, we consider a different branching random walk model, leading to different dynamics (see remarks at the end of this section). We consider a population of abstract particles categorized into a denumerable quantity of types, denoted by $j = 0, 1, 2, \ldots$ and evolving according to the following rules:

1. The lifespans of all particles are iid exponential random variables with mean $1/\lambda$.
2. At the moment of death, a particle of type $j \geq 1$ produces two progeny particles each belonging to type $j + 1$ with probability b, to type $j - 1$ with probability d, and to type j with probability $1 - b - d$. However, a particle of type $j = 0$ produces two progeny of type 0.
3. The process is initiated at time $t = 0$ by a single particle of given type $i > 0$.

We consider the infinite vector $\mathbf{Z}(t) = (Z_0(t), Z_1(t), \ldots)$, where $Z_j(t)$ is the number of particles of type j at time t.

The main results obtained are:

- Exact expressions for the expectations of the process, in the terms of modified Bessel functions
- Asymptotic expressions for the expectations, exponential modified by negative power terms

The distribution of $\mathbf{Z}(t)$ is determined by the probability generating functions (*pgf*'s). Denote $F_i(\mathbf{s}, t)$, the pgf of the infinite vector $\left(Z_0^{(i)}(t), Z_1^{(i)}(t), \ldots\right)$ of particle counts at time t, given that at time $t = 0$, there was exactly one particle of type i. As detailed in Kimmel and Stivers (1994), the pgf's satisfy the following infinite system of ordinary differential equations:

$$\begin{aligned} \frac{\partial F_0(\mathbf{s}, t)}{\partial t} &= \lambda[F_0^2(\mathbf{s}, t) - F_0(\mathbf{s}, t)], \\ \frac{\partial F_i(\mathbf{s}, t)}{\partial t} &= \lambda[f(F_{i-1}(\mathbf{s}, t), F_i(\mathbf{s}, t), F_{i+1}(\mathbf{s}, t)) - F_i(\mathbf{s}, t)] i \geq 1, \end{aligned} \tag{7.6}$$

where $f(s_{i-1}, s_i, s_{i+1}) = ds_{i-1}^2 + (1 - b - d)s_i^2 + bs_{i+1}^2$. The initial condition is $F_i(\mathbf{s}, 0) = s_i$.

Denote $M_{ij}(t)$, the mean number of particles of type j at time t, generated by a process starting with a single particle of type i at $t = 0$. Differentiating Eq. (7.6) with respect to s_j, we obtain

$$\frac{d}{dt} M_{ij}(t) = \lambda \left[2d M_{i-1,j}(t) + [1 - 2(b + d)] M_{ij}(t) + 2b M_{i+1,j}(t)\right], i \geq 1 \tag{7.7}$$

and $M_{0j} = \mathrm{e}^{\lambda t} \delta_{0j}$. Equation (7.7) is a system of linear differential equations.

One way to solve Eq. (7.7) is to construct a generating function of the M_{ij}'s,

$$\mathcal{M}^j(u, t) = \sum_{i \geq 0} u^i M_{ij}(t), j \geq 0.$$

Proceeding from the definition of $\mathcal{M}^j$, we obtain from Eq. (7.7)

$$\frac{d}{dt}\mathcal{M}^j(u,t) = \left[2(b+d) - \frac{2b}{u}\right]\lambda e^{\lambda t}\delta_{0j} + \lambda\left[2du + [1-2(b+d)] + \frac{2b}{u}\right]\mathcal{M}^j(u,t) - 2\lambda b M_{1j}(t).$$

If $j \neq 0$, then $\delta_{0j}(t) = 0$, so

$$\frac{d}{dt}\mathcal{M}^j(u,t) = A(u)\mathcal{M}^j(u,t) - 2b\lambda M_{1j}(t),\ j \geq 1 \tag{7.8}$$

where

$$A(u) = \lambda\left[2du + 1 - 2(b+d) + \frac{2b}{u}\right].$$

Denoting the Laplace transform of $\mathcal{M}$ by $\hat{\mathcal{M}}$, we transform Eq. (7.8) with respect to t (Doetsch 1974):

$$p\hat{\mathcal{M}}^j(u,p) - \mathcal{M}^j(u,0) = A(u)\hat{\mathcal{M}}^j(u,p) - 2b\lambda\hat{M}_{1j}(p).$$

Clearly, $M_{ij}(0) = \delta_{ij}$, so $\mathcal{M}^j(u,0) = u^j$. Therefore, we obtain

$$\hat{\mathcal{M}}^j(u,p) = \frac{-u(u^j - 2b\lambda\hat{M}_{1j}(p))}{2d\lambda u^2 + ([1-2(b+d)]\lambda - p)u + 2b\lambda}. \tag{7.9}$$

$\hat{\mathcal{M}}^j(u,p)$ is analytic in p for any $u \in [0,1)$. Therefore, if $\hat{u} \neq 0$ solves $u[A(u) - p] = 0$, then $\hat{u}$ must also be a root of $u^j - 2b\lambda\hat{M}_{1j}(p)$, which implies that $\hat{M}_{1j}(p) = \hat{u}^j/(2b\lambda)$. The roots of the denominator are:

$$\hat{u}_i = \frac{p - \lambda[1-2(b+d)] + (-1)^i\sqrt{(\lambda[1-2(b+d)] - p)^2 - 16bd\lambda^2}}{4d\lambda},\ i = 1,2. \tag{7.10}$$

Substituting $\hat{u} = \hat{u}_1$ into $\hat{M}_{1j}(p) = \hat{u}^j/(2b\lambda)$, we obtain $\lim_{p\to\infty}\hat{M}_{1j}(p) = 0$ (p real). The other root yields $\lim_{p\to\infty}\hat{M}_{1j}(p) = \infty$, inconsistent with the properties of the Laplace transform. Using $\hat{M}_{1j}(p) = \hat{u}_1^j/(2b\lambda)$, we obtain

$$\hat{M}_{1j}(p) = \frac{1}{2\lambda b}\hat{g}\left(\frac{p}{4\lambda d} - \frac{\lambda[1-2(b+d)]}{4\lambda d}\right), \tag{7.11}$$

where

$$\hat{g}(x) = \left\{x - \sqrt{x^2 - \frac{b}{d}}\right\}^j.$$

The counter image of $\hat{M}_{1j}(p)$ is

$$M_{1j}(t) = \frac{j e^{\lambda[1-2(b+d)]t}}{2b\lambda t} \left(\sqrt{\frac{b}{d}}\right)^j I_j(4\sqrt{bd}\lambda t), \tag{7.12}$$

where $I_j(z)$ is the modified Bessel function of order j (Abramowitz and Stegun 1958).

The following theorem describing the asymptotic behavior of $M_{1j}(t)$ was proved in Kimmel and Stivers (1994):

Theorem 7.2 *Suppose that $b < d$. Then,*

$$M_{1j}(t) \sim K_j \frac{e^{\lambda[1-2(\sqrt{b}-\sqrt{d})^2]t}}{t^{3/2}}, \text{ as } t \to \infty,$$

where

$$K_j = \frac{j\left(\sqrt{b/d}\right)^j}{4\lambda^{3/2}\sqrt{2\pi}b(bd)^{1/4}}, \ j \geq 1,$$

and

$$\sum_{j\geq 1} M_{1j}(t) \sim K_S \frac{e^{\lambda[1-2(\sqrt{b}-\sqrt{d})^2]t}}{t^{3/2}}, \text{ as } t \to \infty,$$

where

$$K_S = \frac{d\sqrt{\pi}}{4\lambda^{3/2}\sqrt{2\pi}(bd)^{1/4}(\sqrt{b}-\sqrt{d})^2}.$$

Moreover,

$$\sum_{j\geq 1} \frac{K_j}{K_S} = 1.$$

The consequence of Theorem 7.2 is that this branching random walk exhibits a property known as quasistationarity (mentioned in the context of Yaglom Theorem 3.6). We see that the entire population $\sum_{j\geq 0} M_{1j}(t)$ grows as $\exp(\lambda t)$. Since $\sum_{j\geq 1} M_{1j}(t)$ grows only as $t^{-3/2}e^{\lambda[1-2(\sqrt{b}-\sqrt{d})^2]t}$, this means that $M_{10}(t)$ grows as $\exp(\lambda t)$, i.e., that type 0 is absorbing in the sense that $M_{10}(t)/\sum_{j\geq 0} M_{1j}(t) \to 1$, as $t \to \infty$. However, $M_{1i}(t)/\sum_{j\geq 1} M_{1j}(t) \to K_i/K_S$, i.e., the distribution of types conditional on nonabsorption tends to a limit (i.e., it reaches the quasistationary distribution). This quasistationary behavior of the random walk with an absorbing barrier is similar to that exhibited by the process of division-within-division (Sect. 7.6). In the next section, Theorem 7.2 will be applied to a model of unstable gene amplification, which may be considered an extension of the model of Sect. 3.7.

This model is a time-continuous generalization of the random walk model from Kimmel and Axelrod (1990). No particular mechanism of gene amplification is assumed. It is only postulated that from one generation to another the number of gene copies on extrachromosomal elements may double or half. This model is based on the so-called quantitative shift model described by Peterson (1984).

Hypotheses:

1. The lifespans of cells are independent random variables distributed exponentially with mean $1/\lambda$.
2. (a) Progeny of a cell having at least two gene copies may have twice as many gene copies per cell, the same number of gene copies per cell, or half as many gene copies per cell, with respective probabilities b, $1 - b - d$, and d.
   (b) Progeny of a cell having a single copy of the gene may have two gene copies per cell, one gene copy per cell, or *no* gene copies per cell, with respective probabilities b, $1 - b - d$, and d.
   (c) Progeny of a cell having no gene copies will also have no gene copies.

Constants b and d are the probabilities of gene *amplification* and *deamplification*. The asymptotic results of Theorem 7.2 apply directly if the following definition is used: A cell belongs to type $j \geq 1$ if it contains 2^{j-1} gene copies. A cell belongs to type $j = 0$ if it contains no gene copies.

Kimmel and Stivers (1994) employed this model to estimate probabilities of gene amplification and deamplification in cultured cells. Further analysis can be found, among others, in Bobrowski and Kimmel (1999).

7.5 Application: Stable Gene Amplification

This is a model for a variant of the process of gene amplification different from the one considered in the chapter on the GW process. The previous model accounted for the instability of some amplified genes by their loss from cells during cell division. The loss of these extrachromosomal elements was associated with the lack of centromeres which are found on chromosomes and are required for faithful segregation at cell division.

The model developed here accounts for situations in which gene amplification can be either stable or unstable. It is based on different experimental observations and a more extensive biological model (Windle et al. 1991). It describes the initiation of amplification as the breakage of a piece of a chromosome releasing a fragment containing a gene or genes but not a centromere. Genes on these acentric extrachromosomal fragments may replicate and recombine forming increased numbers of tandemly repeated genes. As in the previous model, the extrachromosomal elements may segregate, although their segregation is not faithful because they do not have the required centromeres.

A new aspect of the model is that it also includes the possibility of stabilization of the number of amplified genes following their reintegration into chromosomes. This

is because reintegration links the amplified genes to the centromeres on chromosomes allowing them to be faithfully segregated at cell division.

From the mathematical viewpoint, this model is an example of a decomposable process. Decomposable processes include a subclass of transient types which are irreversibly lost from the process. Processes limited to the remaining types may behave variously. In this particular application, as we will see, it will reach a limit distribution. A decomposable process cannot be positive regular, as a whole, although the persistent subprocess may be.

7.5.1 Assumptions

The following is the list of model hypotheses (Fig. 7.1):

1. All acentric (extrachromosomal) elements evolve independently of each other.
2. Types:
   (a) Acentric elements containing $i = 1, 2, \ldots$ gene copies
   (b) Chromosomes with one or more sites containing reintegrated elements, each containing $i = 1, 2, \ldots$ gene copies.
3. In each cell generation, three types of events can occur for each acentric extrachromosomal element:
   (a) Element replicates and breaks at a random site, and the pieces segregate.
   (b) Element replicates and does not break.
   (c) Element reintegrates into a chromosome.
4. With probability a, the element with i gene copies replicates and yields a single element with $2i$ gene copies, and then breaks at a random site producing two pieces with lengths j and $2i - j$, where $j = 1, \ldots, 2i - 1$. The probability of breakage at each site is the same, and equal to $1/(2i - 1)$. The pieces segregate so that they both pass to the same progeny cell with probability α, and pass to different progeny cells with probability $1 - \alpha$.
5. With probability b, the element with i gene copies replicates to yield a single element with $2i$ gene copies, but it does not break. It then passes with probability $1/2$ to one of the two progeny.
6. With probability c, the element containing i copies of the gene is integrated into a chromosome with a centromere and then replicates and segregates with the chromosome. This results in progeny cells with equal number of gene copies. No further breakage, nor increase or decrease in gene copy number occurs at this site at subsequent cell divisions. The probability of reintegration is equal to $c = 1 - (a + b)$.

Initial conditions. At the beginning of the process, a single cell contains a single extrachromosomal element with one gene copy, $i = 1$. It is understood that this element was formed in the past by deletion of one copy of a chromosomal gene in a founder cell.

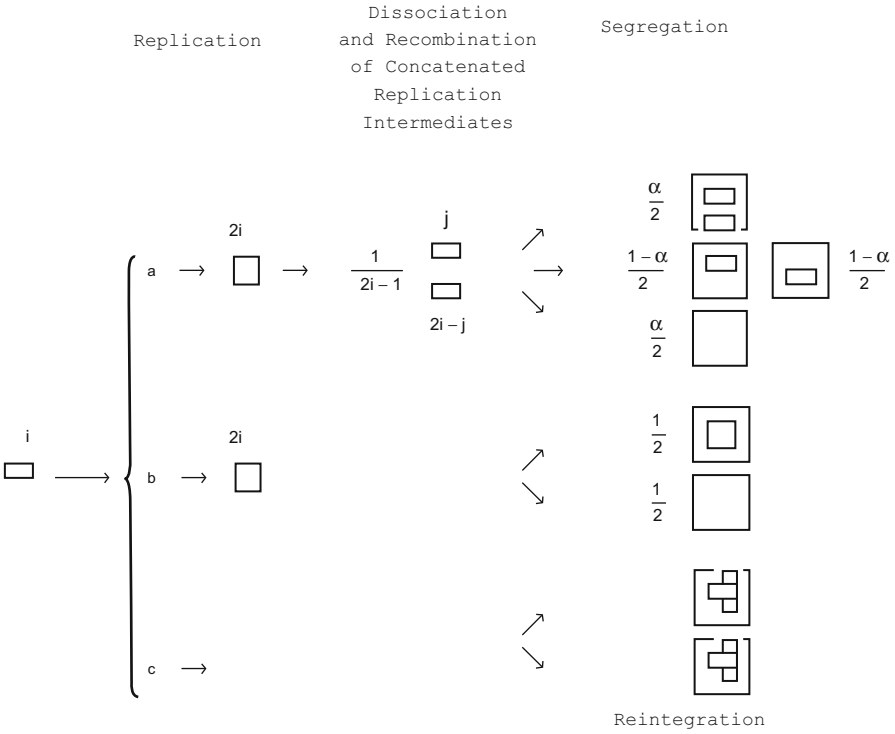


Fig. 7.1 Schematic representation of the events in the gene amplification model. (Source: Kimmel et al. 1992)

Remark

1. Breakage can be understood as imperfect resolution of replicated DNA.
2. If breakage occurs, a randomly chosen progeny cell will contain both pieces with probability $\alpha/2$, no piece with probability $\alpha/2$, a single piece of size j with probability $(1-\alpha)/2$, and a single piece of size $2i-j$ with probability $(1-\alpha)/2$.

Consequences

1. In successive cell generations cells with no gene copies are produced. These are killed by a selective agent.
2. Among cells with at least one gene copy, there will be an initial increase in number of extrachromosomal elements per cell, and number of gene copies per extrachromosomal element. Subsequently, as the acentric elements become reintegrated, their number per cell will decrease and the proportion of cells with stably integrated copies will increase (as observed).
3. Eventual consequence will be a population of cells containing only one or more integrated elements with a distribution of gene copy numbers. It is possible to compute this distribution.

7.5.2 Pgf's and Expectations

The process includes an infinite spectrum of chromosomal and extrachromosomal elements with $1, 2, \ldots$, copies of the gene. We consider a randomly selected line of descent of cells. We define the following random variables:

- X_n^i, the number of extrachromosomal elements with i copies of the gene, in the nth cell generation.
- Y_n^i, the number of elements reintegrated into chromosomes, with i copies of the gene, in the nth cell generation.

The sequence $\{\{(X_n^1, Y_n^1), (X_n^2, Y_n^2), \ldots\}, n = 0, 1, 2, \ldots\}$, is a *multitype GW process with a denumerable infinite number of particle types*.

Let us consider the aggregated process $\{(X_n, Y_n),\ n = 0, 1, 2, \ldots\}$, where

$$X_n = \sum_{i=1}^{\infty} X_n^i,\ Y_n = \sum_{i=1}^{\infty} Y_n^i,$$

are the total number of the extrachromosomal elements and elements reintegrated into chromosomes, in generation n. Following the rules of the process (see Fig. 7.1), the pgf of the number of progeny of an extrachromosomal element is equal to

$$f^1(s_1, s_2) = \frac{a\alpha}{2} s_1^2 + \left[a(1-\alpha) + \frac{b}{2}\right] s_1 + \frac{a\alpha + b}{2} + cs_2. \tag{7.13}$$

The coefficient of the quadratic term reflects the fact that two extrachromosomal elements are produced from a single one only if breakage occurs (wp a), both elements segregate into one progeny (wp α), and this progeny belongs to the lineage followed (wp $\frac{1}{2}$). The remaining coefficients are derived analogously.

The pgf of the number of progeny of a reintegrated element is simply $f^2(s_2) = s_2$, since such element is stable. Let us denote by $f_n^1(s_1, s_2)$ the joint pgf of $\{(X_n, Y_n)|(X_0, Y_0) = (1, 0)\}$. The following relationship can be derived using the backward approach as in Sect. 3.2:

$$f_{n+1}^1(s_1, s_2) = \frac{a\alpha}{2}[f_n^1(s_1, s_2)]^2 + \left[a(1-\alpha) + \frac{b}{2}\right] f_n^1(s_1, s_2) + \frac{a\alpha + b}{2} + cs_2. \tag{7.14}$$

This equation provides a recursive procedure for finding distributions of the process.

The process tends to a nontrivial limit with probability 1,

$$(X_n, Y_n) \longrightarrow (0, Y_\infty),\ \text{wp } 1,\ \text{as } n \to \infty. \tag{7.15}$$

To demonstrate it, let us first note that $Y_{n+1}(\omega) \geq Y_n(\omega)$ which yields the almost sure convergence of Y_n, with the limit being possibly an improper random variable.

However, passing to infinity with n in Eq. (7.14) yields the following quadratic equation for the pgf of Y_∞:

$$\frac{a\alpha}{2}[f_\infty^1(1,s_2)]^2+\left[a(1-\alpha)+\frac{b}{2}-1\right]f_\infty^1(1,s_2)+\frac{a\alpha+b}{2}+cs_2=0. \quad (7.16)$$

The pgf solution of (7.16) verifies $f_\infty^1(1,s_2)|_{s_2=1}=1$, which means that Y_∞ is a proper random variable. On the other hand, if we set $s_2=1$ in Eq. (7.13), we see that $\{X_n\}$ is a subcritical GW process, which yields $\mathrm{P}\{\lim_{n\to\infty}X_n=0\}=1$. This completes the proof of (7.15).

Let us note that in an experimental setting, only cells with nonzero number of elements (extrachromosomal or reintegrated) are observed, since only these cells survive under drug selection pressure. Therefore, all distributions should be conditional on nonextinction of the process, i.e., on the event $\{(X_n,Y_n)\neq(0,0)\}$. In particular, the conditional probability that in generation n, extrachromosomal elements are still present in the lineage, is

$$r_n=\mathrm{P}\{X_n>0|(X_n,Y_n)\neq(0,0)\}=\frac{1-f_n^1(0,1)}{1-f_n^1(0,0)}. \quad (7.17)$$

Let us consider the expectations of the complete process $\{\{(X_n^1,Y_n^1),(X_n^2,Y_n^2),\dots\},n=0,1,2,\dots\}$,

$$\mu_n^i=\mathrm{E}(X_n^i),\ \ \nu_n^i=\mathrm{E}(Y_n^i),\ \ i\geq 1,\ n\geq 0. \quad (7.18)$$

It is not difficult to verify that the infinite vector $\mu_n=\{\mu_n^1,\mu_n^2,\dots\}$ satisfies the recurrence

$$\mu_{n+1}=\mu_n\mathcal{M},\ n\geq 0;\ \ \mu_0^i=\delta_{1i}, \quad (7.19)$$

where $\mathcal{M}$ is an infinite matrix of the form

$$\mathcal{M}=\begin{pmatrix} a & b/2 & 0 & & & & & & & \mathbf{0}\\ a/3 & a/3 & a/3 & b/2 & 0 & & & & & \\ a/5 & a/5 & a/5 & a/5 & a/5 & b/2 & 0 & & & \\ a/7 & a/7 & a/7 & a/7 & a/7 & a/7 & a/7 & b/2 & 0 & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}. \quad (7.20)$$

The expectations of Y_n^i satisfy

$$\nu_{n+1}^i=\nu_n^i+c\mu_n^i,\ n\geq 0,\nu_0^i=0. \quad (7.21)$$

Let us note that $\nu_{n+1}^i\geq\nu_n^i$. Using this and an analysis involving (7.16), we obtain that $\lim_{n\to\infty}\sum_{i=1}^\infty\nu_n^i=\lim_{n\to\infty}\mathrm{E}(Y_n)<\infty$. Therefore, $\lim_{n\to\infty}\nu_n^i$ exist and are finite ($i\geq 1$).

The expectations $\mu_n^i = \mathrm{E}(X_n^i)$ and $\nu_n^i = \mathrm{E}(Y_n^i)$, properly normed, can be understood as distributions of the sizes of extrachromosomal and reintegrated elements in the lineage. To calculate these expectations, it is convenient to introduce the generating functions

$$M_n(z) = \sum_{i=1}^{\infty} \mu_n^i z^i,\ N_n(z) = \sum_{i=1}^{\infty} \nu_n^i z^i,\ z \in [0,1].$$

Equations (7.19) and (7.21) yield the following relationships for the generating function:

$$M_{n+1}(z) = \frac{b}{2} M_n(z^2) + \frac{az}{1-z} \int_z^1 \frac{M_n(u^2)}{u^2} \mathrm{d}u,\ n \geq 0,$$

where $M_0(z) = z$, and

$$N_n(z) = c \sum_{k=0}^{n-1} M_k(z).$$

After carrying out differentiations with respect to z in the first of the relationships above, we obtain that the mean size of the extrachromosomal elements in generation n is equal to

$$\frac{M_n'(1-)}{M_n(1)} = \left(\frac{b+a}{\frac{b}{2}+a} \right)^n,$$

which tends to ∞, as $n \to \infty$. The expected size of reintegrated elements has a finite limit

$$\frac{N_\infty'(1-)}{N_\infty(1)} = \frac{1-a-\frac{b}{2}}{1-a-b}.$$

7.5.3 *Model Versus Data*

Parameters of the model for a single experimental system can be deduced based on experiments by Geoffrey Wahl and colleagues of the Salk Institute (Windle et al. 1991). From his results, it is possible to estimate the following quantities:

1. The fraction of nonextinct cells still containing extrachromosomal elements after 9 generations ($r_9 \sim 0.39$).
2. The fraction of nonextinct cells still containing extrachromosomal elements after 35 generations ($r_{35} \sim 0.02$, highly inaccurate).
3. The fraction of nonextinct cells with 1–2 elements extrachromosomal and/or reintegrated (as opposed to cells containing $\geq$ 3 elements), after 9 generations ($p_{12} \sim 0.54$).

The theoretical relationships (7.17) and (7.14), with parameter values

$$\alpha = 1,\ a = 0.92,\ b = 0.035,$$

yield

$$r_9 = 0.39,\ r_{35} = 0.05,\ p_{12} = 0.63,$$

in approximate agreement with the experiment.

The expected size of reintegrated elements predicted by the model is rather low, equal to $4/3$. However, this is based on the assumption that the initial extrachromosomal element in generation 0 is of a unit size.

7.6 Application: Quasistationarity in a Branching Model of Division-Within-Division

Branching-within-branching occurs in various settings in cell and molecular biology. Examples include tightly regulated phenomena like replication of chromosomal DNA, and also processes in which the number of objects produced in each biological cell is a random variable. These are gene amplification in cancer cells, plasmid dynamics in bacteria, and proliferation of viral particles in host cells.

The general motivating idea is stability arising from selection superimposed on a random mechanism. We consider a set of large particles (biological cells), following a binary fission process. Each of the large particles is born containing a number of small particles (genes, viruses, organelles), which multiply or decay during the large particle's lifetime. The arising population of small particles is then split between the two progeny of the large particle and the process continues in each of them.

Let us suppose that the presence of at least one small particle is necessary to ensure the viability of the large particle. This can be due to a selection factor existing in the environment. One example of such selection factor is a cytotoxic drug, which eliminates cells (large particles) devoid of resistance genes (small particles), as in the gene amplification model of Sect. 7.4. We are interested in the behavior of the population of large particles surviving selection, i.e., large particles having at least one small particle inside.

We show that if the smaller particles follow a subcritical process, the number of smaller particles contained in a nonextinct large particles tends to a limit distribution. The result, in its present form (Kimmel 1997), depends on several detailed hypotheses, but these can be relaxed.

A similar idea has been explored in the paper by Bühler (1992), where a population (whose individuals are called cells) develops according to a GW branching process. The cells are living in separate groups called subunits. Whenever a subunit becomes "too big," it splits into two or more new subunits. The number and the sizes of the new subunits are determined independently from the behavior of all the cells in the other subunits and of the prehistory of the process, the only restriction being the

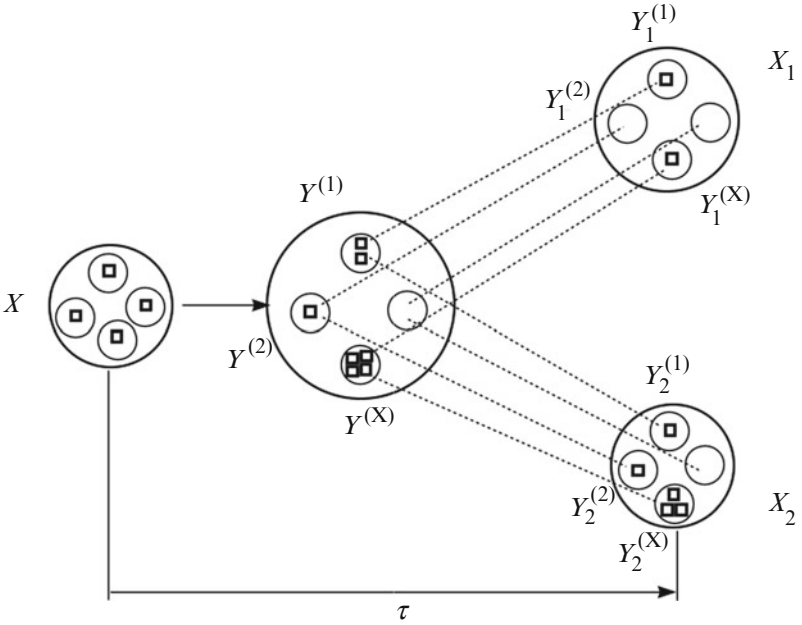


Fig. 7.2 Large particle containing X small particles lives for a random time τ exponentially distributed with parameter λ and then splits into two progeny. During its lifetime each of the X small particles proliferate producing correspondingly $Y^{(1)}, Y^{(2)}, \ldots, Y^{(X)}$, small particles. Each of these $Y^{(k)}$'s is split independently among the two progeny of the large particle, so that large progeny 1 and 2 receive $\sum_{k=1}^{X} Y_1^{(k)}$ and $\sum_{k=1}^{X} Y_2^{(k)}$ small particles, respectively. The joint distributions of the pairs $(Y_1^{(k)}, Y_2^{(k)})$ are identical and symmetric. (Source: Arino et al. 1995)

obvious one that the total number of cells in "daughter" subunits be the number of daughter cells from the cells of the "mother" subunit and that no empty subunits be formed. The work was stimulated by the model of growth of intestinal crypts by Loeffler and Grosmann (1991).

7.6.1 Definition of the Process

Rules (schematically depicted in Fig. 7.2):

1. The population of large particles evolves according to a binary-fission time-continuous Markov branching process (Yule process), i.e., each particle lives for a random time τ, exponential with parameter λ, and then splits into two progeny, each of which independently follows the same scenario.
2. Each large particle contains X small particles at its birth. Each of these proliferates producing

$$Y^{(1)}, Y^{(2)}, \ldots, Y^{(X)}, \tag{7.22}$$

small particle progeny at the end of the large particle's lifetime.

3. Each of the $Y^{(k)}$ progeny of the initial kth small particle is independently split between the progeny of the large particle, so that large progeny 1 and 2 receive correspondingly $Y_1^{(k)}$ and $Y_2^{(k)}$ small progeny. The joint distributions of the pairs $(Y_1^{(k)}, Y_2^{(k)})$ are identical, independent for all (k), and symmetric in $Y_1^{(k)}$ and $Y_2^{(k)}$. They are described by the joint pgf

$$f_{12}(s_1, s_2) = \mathrm{E}[s_1^{Y_1^{(1)}} s_2^{Y_2^{(1)}}]. \tag{7.23}$$

4. As a result, each of the large progeny receives the total of

$$X_1 = \sum_{k=1}^{X} Y_1^{(k)} \text{ and } X_2 = \sum_{k=1}^{X} Y_2^{(k)} \tag{7.24}$$

small progeny particles.

The resulting branching process can be described as a Markov time-continuous process with denumerable infinity of types of large particles. The large particle is *of type* i if it contains i copies of small particles at its birth. Let us denote the vector of counts of large particles of all types at time t, by $Z(t) = [Z_0(t), Z_1(t), Z_2(t), \ldots]$, and the infinite matrix of expected values $M(t) = [M_{ij}(t)]$ by $M_{ij}(t) = \mathrm{E}[Z_j(t)|Z_i(0) = 1,\ Z_k(0) = 0,\ k \neq i]$.

Let us define coefficients $a_{nm}(i)$ using the expansion of the pgf of the sums in Eq. (7.24) given $X = i$,

$$[f_{12}(s_1, s_2)]^i = \sum_{n,m \geq 0} a_{nm}(i) s_1^n s_2^m, \tag{7.25}$$

$a_{nm}(i)$ is equal to the probability that among the progeny of the i small particles present at birth of the large particle, n will end in large progeny 1, and m will end in large progeny 2.

The expected value equations are obtained in the way analogous as in Sect. 4.2.1 (Kimmel 1997):

$$\frac{d}{dt} M(t) = \lambda(2A - I)M(t),\ M(0) = I, \tag{7.26}$$

where $A = [A_{ij}] = [a_j(i)]$ is the matrix of coefficients of the marginal pgf of X_1 given $X = i$

$$[f(s_1)]^i = [f_{12}(s_1, 1)]^i = \sum_{j,l \geq 0} a_{jl}(i) s_1^j = \sum_{j \geq 0} a_j(i) s_1^j, \tag{7.27}$$

and I is the infinite identity matrix. $a_j(i)$ is equal to the probability that of the i small particles present in the large particle at its birth, j will end in large progeny 1 (or in large progeny 2).

Equation (7.26) can be explicitly solved using the Laplace transform techniques. The solution can be expressed in the form of generating function

$$M_k(u,t) = \sum_{l\geq 0} M_{kl}(t)u^l,\ u \in [0,1]. \tag{7.28}$$

We obtain

$$M_k(u,t) = \sum_{j\geq 0} \frac{(2\lambda t)^j}{j!}[f_j(u)]^k \mathrm{e}^{-\lambda t},\ k \geq 0. \tag{7.29}$$

where $f_j(u)$ is the jth iterate of the marginal pgf of $Y_1^{(1)}$.

7.6.2 Quasistationarity

We begin with stating several facts concerning the GW process with progeny pgf $f(u)$ (see Sect. 3.5.2).

If $f'(1-) < 1$ (the subcritical case) then as $j \to \infty$,

$$\frac{f_j(u) - f_j(0)}{1 - f_j(0)} \to \mathcal{B}(u), \tag{7.30}$$

i.e., conditionally on nonextinction, the process tends to a limit distribution, with pgf $\mathcal{B}(u)$ such that $\mathcal{B}(0) = 0$, $\mathcal{B}(1) = 1$ (c.f. Athreya and Ney 2004, Corollary I.8.1). This behavior is known as quasistationarity. Moreover, as $j \to \infty$

$$f_j(u) - 1 \ \sim\ \rho^j Q(u), \tag{7.31}$$

where $\rho = f'(1-)$ and the function $Q(u)$ satisfies

$$\frac{Q(0) - Q(u)}{Q(0)} = \mathcal{B}(u), \tag{7.32}$$

with $Q(1) = 0$, $Q'(1-) = 1$, $Q(u) \leq 0$, and $Q(u)$ increasing for $u \in [0,1]$.

Functions $\mathcal{B}(u)$ and $Q(u)$ are unique solutions of certain functional equations (Sect. 3.5.2) . The following results are proved in Kimmel (1997):

Theorem 7.3 *Let us consider the process defined in* Sect. 7.6.1 *started by a large ancestor of type k and let* $\rho = f'(1-) < 1$. *Then, as* $t \to \infty$,

$$\mathrm{e}^{\lambda t} - M_k(u,t) \ \sim\ -kQ(u)\mathrm{e}^{(2\rho-1)\lambda t}, \tag{7.33}$$

for all $k \geq 1$.

Corollary 7.1 *The expected frequencies* $\{\mu_{kl}(t),\ l \geq 1\}$ *of large particles of type l among the particles of nonzero type tend, as* $t \to \infty$, *to a limit distribution independent of k, characterized by the pgf* $\mathcal{B}(u)$.

7.6.3 Gene Amplification

The process considered here serves as another model of gene amplification. It is a direct generalization of GW process models of Sect. 3.7.2. Let us assume that large particles are cells and the small ones are copies of the gene conferring drug resistance located on extrachromosomal elements. Cells without any copies of the gene are eliminated by the drug (the selective agent). We accept the following specific hypotheses, similar to those in Kimmel and Axelrod (1990) and Kimmel and Stivers (1994):

- During cell's lifetime each extrachromosomal copy of the gene is successfully replicated with probability β, less than 1.
- The resulting two copies are segregated to the same progeny cell with probability α and to two different progeny cells with probability $1-\alpha$. α may be called *the probability of cosegregation* and has been shown to be ≈ 0.9 in one cell system.

The above hypotheses yield

$$f_{12}(s_1, s_2) = \beta\left[(1-\alpha)s_1 s_2 + \frac{\alpha}{2}\left(s_1^2 + s_2^2\right)\right] + (1-\beta), \tag{7.34}$$

and

$$f(u) = \frac{\beta\alpha}{2}u^2 + \beta(1-\alpha)u + \left(\frac{\beta\alpha}{2} + 1 - \beta\right), \tag{7.35}$$

with $\rho = f'(1-) = \beta < 1$. Therefore our theorem and its corollary apply.

Qualitatively, all the experimental observations above are explained by our results: The stable quasistationary distribution of gene copy count is predicted by Corollary 7.1.

If the type 0 cells are not removed by the drug, then the theorem proves they dominate the population. Indeed by the theorem the resistant cells grow as

$$\sum_{l\geq 1} M_{kl}(t) = M_k(1,t) - M_k(0,t) \sim -kQ(0)\mathrm{e}^{(2\rho-1)\lambda t},\ \rho < 1, \tag{7.36}$$

while the entire population grows as $\mathrm{e}^{\lambda t}$.

If $\rho > 1/2$, then in the presence of the drug, resistant cells grow as $\mathrm{e}^{(2\rho-1)\lambda t}$, i.e., exponentially but slower than in the nonselective conditions.

7.7 Application: Mathematical Modeling of the Loss of Telomere Sequences

7.7.1 Stochastic Model

Telomeres are structures at the ends of chromosomes. They consist of repeated DNA sequences which play a role in replication of the ends of DNA, and in preventing

the ends of chromosomes from sticking together. The number of repeat sequences of the telomeres is variable, and on the average, declines with the increasing number of divisions of normal cells in culture, and of somatic cells in organisms. Reviews on the biology of telomeres include Blackburn (1991), Greider and Blackburn (1996), Greider (1996), Zakian (1995, 1996).

Cellular senescence, the loss of capacity to proliferate, seems to be associated with the inability to maintain a minimum number of telomere sequences. In contrast, immortalized cells such as cancer cells, seem to be able to maintain a low but effective number of telomere sequences. The first researcher who noticed the relationship between telomere endings and cell senescence was Olovnikov (1973). He correctly attributed this loss to the end-replication problem, which arises because of the inability of the DNA polymerase to replicate the downstream end of the DNA molecule. The effect is that each successive DNA replication results in a copy, which is shorter at one end.

Our model (Arino et al. 1995) describes shortening of telomeres by incomplete replication. The two uses of the model are predictions of (1) the expected telomere length and (2) of the fraction of viable cells in aging cell populations. For these purposes, it is first necessary to describe the dynamics of telomere loss from a single chromosome. For simplicity, we proceed as if the process of telomere loss ended when all the telomere *deletion units*, each containing possibly more than a single DNA repeat, are lost. The same mathematics applies to telomere loss until a particular *checkpoint* is encountered.

A chromosome is an entity with a centromere, while a chromatid is a double helix composed of two single strands of DNA. In G_1 phase of the cell cycle, before DNA replication, a chromosome is composed of one chromatid, while in G_2 and M phases, after DNA replication, a chromosome is composed of two chromatids. Levy et al. (1992) described telomere loss in terms of what happens to single DNA strands in G_1. We follow that description. Figure 7.3 depicts the scheme of deletion and segregation of telomere sequences on chromosome ends. It can be summarized mathematically as follows:

- Each chromatid is composed of two strands named *upper* or $5' \to 3'$, and *lower* or $3' \to 5'$, each of which has two ends named *left* and *right*. The numbers of telomeric deletion units on both ends of both strands are symbolically represented by quadruples of the form $(a, b; c, d)$, where a and c correspond to the left and right ends of the upper strand, while b and d correspond to the left and right ends of the lower strand. The only important combinations of a, b, c, and d are of the form $(n-1, n; m, m)$ or $(n, n; m, m-1)$, since they always arise after a single replication round (details not shown).
- Cells containing chromatids described by the quadruple $(n-1, n; m, m)$ give birth to two progeny containing chromatids of the types $(n-1, n; m, m)$ and $(n-1, n-1; m, m-1)$, respectively. This transition rule as well as the dual rule for the other admissible type are depicted symbolically below. Let us note that one progeny is always of the same type as the parent cell, while the other is missing two sequences, each on a different end of a different strand:

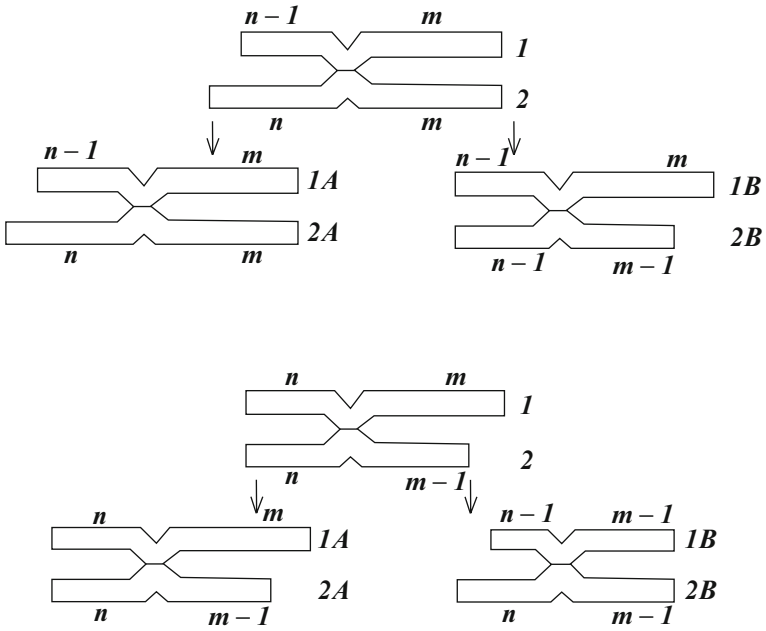


Fig. 7.3 Transition rules for deletion and segregation of telomere ends on a chromosome in G_1. DNA strands 1 and 2 replicate and segregate into different daughter cells A and B, resulting in chromatids (1A, 2A) and (1B, 2B), respectively. Due to the end-replication problem, one DNA strand on each of the newly created chromatids contains additional deletion at its right or left end, depending on its orientation and presence of the deletion on the corresponding strand of the mother chromatid. For additional explanations see Levy et al. (1992). (Source: Kimmel and Axelrod 1991)

$$(n-1,n;m,m) \begin{cases} \to & (n-1,n;m,m) \\ \to & (n-1,n-1;m,m-1) \end{cases}$$

$$(n,n;m,m-1) \begin{cases} \to & (n,n;m,m-1) \\ \to & (n-1,n;m-1,m-1). \end{cases}$$

- Proliferation ends when the telomere ends become short enough. Without a loss of generality, we assume that cells of the types $(n-1,n;0,0)$ and $(0,0;m,m-1)$ have a single progeny of the type identical to that of the parent cell, i.e.:

$$(n-1,n;0,0) \quad \to \quad (n-1,n;0,0)$$

$$(0,0;m,m-1) \quad \to \quad (0,0;m,m-1).$$

If we renumber states in such way that index $k = 0, 1, \ldots,$ is equal to the sum of numbers of deletion units on the left ends of the upper and lower strand, and index $l = 0, 1, \ldots,$ is equal to the sum of numbers of deletion units on the right ends of the upper and lower strand:

$$k = \begin{cases} & 2n, \quad \text{if } (n, n; m, m-1) \text{ occurs,} \\ \text{or} & \\ & 2n-1, \quad \text{if } (n-1, n; m, m) \text{ occurs,} \end{cases} \tag{7.37}$$

$$l = \begin{cases} & 2m, \quad \text{if } (n-1, n; m, m) \text{ occurs,} \\ \text{or} & \\ & 2m-1, \quad \text{if } (n, n; m, m-1) \text{ occurs,} \end{cases} \tag{7.38}$$

then the admissible transitions become:

$$(k,l) \quad \begin{cases} \to & (k,l) \\ \to & (k-1, l-1) \end{cases} \tag{7.39}$$

$$(k,0) \quad \to \quad (k,0) \tag{7.40}$$

$$(0,l) \quad \to \quad (0,l). \tag{7.41}$$

In the array (k, l), where k and l are nonnegative integers, the admissible transitions belong to disjoint paths which can be numbered by $k - l$ (path number assuming values from $-\infty$ through ∞). Each of these paths can be treated separately. The state number within each path can be taken as $i = \min(k, l)$. Biologically, it is the number of deletion units on the shorter, and therefore limiting, end. Now the transitions have the form

$$i \quad \begin{cases} \to & i \\ \to & i-1 \end{cases} \tag{7.42}$$

$$0 \quad \to \quad 0. \tag{7.43}$$

7.7.2 Branching Process

Let us assume that lifelengths of cells are *independent identically distributed random variables* with distribution with density $g(t)$ and cumulative distribution $G(t)$. Let us denote

$$X_{ij}(t),\ t \in [0, \infty),\ i, j = 0, 1, \ldots,$$

the family of random variables equal to the number of cells in state j at time t, in the process started at time 0 by a single cell in state i.

Our process can be described as a branching random walk. In our case this means that the type of the progeny object (chromosome) is either identical with the parental type or it is shortened by a single unit. We have,

$$X_{ij}(t) = \begin{cases} X_{ij}(t-\tau) + X_{i-1,j}(t-\tau); & \tau \le t, \\ \delta_{ij}; & \tau > t, \end{cases} \tag{7.44}$$

for all $j = 0, 1, \ldots$, $i = 1, 2, \ldots$ and $t \in [0, \infty)$. The above equation expresses the fact that the number of cells in state j at time t, in a process started at time 0 by a single cell in state i, is either equal to δ_{ij}, if the ancestor cell is still alive, or it is equal to the sum of the numbers of cells in state j at time t in two subprocesses started at time τ (i.e., at the moment of the ancestor's death) by the two progeny of the ancestor, one of which is in state i and the other in state $i-1$. Another equation,

$$X_{0j}(t) = \begin{cases} X_{0j}(t-\tau), & \tau \le t, \\ \delta_{0j}, & \tau > t, \end{cases} \tag{7.45}$$

for all $j = 0, 1, \ldots$ and $t \in [0, \infty)$ expresses the fact that cells in state 0 do not proliferate.

Let

$$M_{ij}(t) = \mathrm{E}[X_{ij}(t)] \tag{7.46}$$

denote the expected count of cells in state j at time t in a process started by an ancestor of type i. We obtain the following equation for the matrix $M(t) = [M_{ij}(t)]$:

$$M(t) = Ag(t) * M(t) + \bar{G}(t)I, \tag{7.47}$$

where $\bar{G}(t) = 1 - G(t)$ and symbol "$*$" denotes convolution of matrix functions on $[0, \infty)$, $g(t) * M(t) = \int_0^t g(t-\tau)M(\tau)\mathrm{d}\tau$, I is the infinite identity matrix and the infinite matrix A has the form,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & & \ddots & \ddots \end{pmatrix}.$$

The solution of this *backward* equation can be represented as an infinite series

$$M = \left[\sum_{k\ge 0} (Ag)^{*k}\right] * \bar{G}I = \bar{G}I * \left[\sum_{k\ge 0} (Ag)^{*k}\right].$$

The second series is the solution of a dual *forward* equation

$$M(t) = M(t) * Ag(t) + \bar{G}(t)I, \tag{7.48}$$

which is equivalent to the system

$$M_{ij}(t) = g(t) * [M_{ij}(t) + M_{i,j+1}(t)] + \delta_{ij}\bar{G}(t),\ j = 0, 1, \ldots,\ t \geq 0, \tag{7.49}$$

which can be examined separately for each ancestor's state i. This would not be possible with the backward system.

7.7.3 Analysis in the Markov Case

If the cell lifelength distributions are exponential, i.e., the density has the form $g(t) = \alpha \exp(-\alpha t)$, the system of convolution Eqs. (7.49) is equivalent to the following infinite system of differential equations:

$$\dot{M}_{ij}(t) = \alpha M_{i,j+1}(t),\ M_{ij}(0) = \delta_{ij},\ j = 0, 1, \ldots,\ t \geq 0. \tag{7.50}$$

This system has an explicit solution,

$$M_{ij}(t) = \begin{cases} \frac{\alpha^{i-j}t^{i-j}}{(i-j)!}, & 0 \leq j \leq i, \\ 0, & j > i. \end{cases} \tag{7.51}$$

Let $M_j(t)$ denote the expected number of cells in state j at time t, if the initial expected counts of cells in states $0, 1, \ldots$ were $M_0(0), M_1(0), \ldots$. Expressions for $M_j(t)$ are obtained by combining solutions of Eq. (7.50).

If the initial cells belong to finitely many different states, so that,

$$M_j(0) = 0,\ j > N, \tag{7.52}$$

then

$$M_j(t) = \sum_{k=j}^{N} M_k(0)\frac{\alpha^{k-j}t^{k-j}}{(k-j)!}. \tag{7.53}$$

We may notice that the polynomial dynamics of the expected values is a consequence of the one-way means of communication between types in the process. This is known as reducibility of the process. Biologically, it is a consequence of the fact that loss of telomere repeats is irreversible.

7.7.4 Model Versus Data

The Markov branching process model was employed to reproduce experimental data on telomere loss. Let us suppose that the number of telomeric repeats in a given chromosome at the time the clonal population growth is initiated ($t = 0$) exceeded the critical (checkpoint) length by d deletion units i.e.,

$$M_j(0) = \delta_{jd} N_0. \tag{7.54}$$

As Levy and coworkers (1992) point out, it is likely that telomeres on different chromosomes differ in their initial number of TTAGGG repeats. Since only the chromosomes with the shortest telomeres are relevant to replicative senescence, only the deletions on the shorter of these chromosomes' two telomeres need to be considered. Suppose there are k such chromosomes with the same critical d, and they segregate independently and that only one critically short telomere is sufficient to signal the cell cycle exit.

We identified two sources of data useful for modeling. One is the paper by Harley and Goldstein (1980) in which fractions $F(d, t)$ of proliferating cells were measured at different times after a clonal culture had been established. These data have been used for modeling by Levy and coworkers (1982; see their Fig. 6).

Another source is the paper by Counter and coworkers (1992) which includes experimental data on the expected telomere lengths (mean number of excess deletion units) $n(t)$.

Our expressions for the expected frequencies of telomere repeat counts on a single chromosome can be combined to yield expressions for $F(d, t)$ and $n(t)$ (details in Arino et al. 1996).

Figure 7.4 depicts the results of modeling of the fraction of viable cells as a function of the number of cell doublings of a clonal culture. Experimental data for two independent cultures of a human fibroblast strain (Harley and Goldstein 1980; after Levy et al. 1992, Fig. 6, modified) are compared to predictions using the Markov branching process model.

The fit has been obtained with parameters $d = 65$ and $k = 40$. Note that the number of chromosomes has to be set equal to $k = 40$ to achieve an acceptable fit, otherwise the decrease in $F(d, t)$ is not sharp enough. This number is not very different from the number of human chromosomes (equal to $2 \times 23 = 46$), which may be taken to mean that all chromosomes have the same critical d-value.

Figure 7.5 depicts the results of modeling the mean length of terminal restriction fragments (TRFs) in function of the number of cell doublings. Experimental data for a number of cultures of normal and transfected cells, up to the crisis time (from Counter et al. 1992) are compared to predictions using the Markov branching process model. The fit has been obtained using parameters $d = 65$ and $k = 40$.

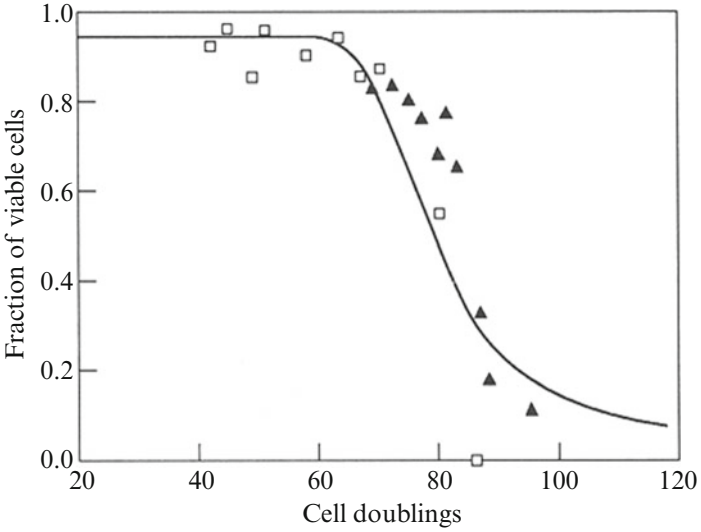


Fig. 7.4 Fraction of viable cells versus the number of cell doublings. Experimental data for two independent cultures of a human fibroblast strain represented by triangles and squares, are compared to predictions using the Markov branching process model (continuous lines), with parameters $d = 65$ and $k = 40$) and with a correction for growth fraction of 0.95. (Source: Arino et al. 1985)

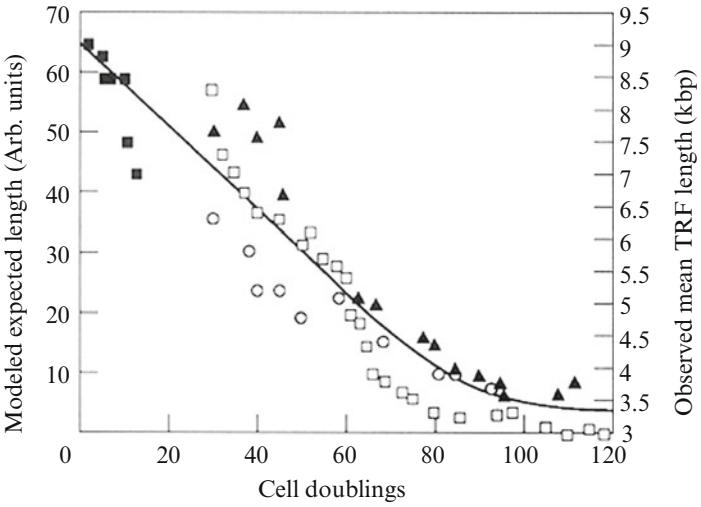


Fig. 7.5 Mean length of terminal restriction fragments (TRF's) versus the number of cell doublings. Experimental data for a number of cultures of normal and transfected cells, up to the crisis time. Different symbols for different experiments, are compared to predictions using the Markov branching process model (continuous line) with parameters $d = 65$ and $k = 40$. (Source: Arino et al. 1985)

7.7.5 *Further Work on Telomere Modeling*

Subsequently, Olofsson and Kimmel (1999) and Olofsson (2000) considered models of telomere shortening involving the possibility of cell death, with the probability of the latter depending on cell type. These models exhibit a variety of limit behaviors, being the consequence of reducibility. The basic tools are the Tauberian theorems for pgf's.

An interesting addition to the literature on stochastic models of telomere dynamics is the paper by Olofsson and Bertuch (2010), which proposes a mechanism for telomere shortening followed by escape into growth (survivorship). They use a model in the form of a general (Jagers) branching process and relate it to data on a population of cells of the yeast *S. cerevisiae* following loss of telomerase. Experimental data indicate that a population of telomerase-deficient cells regains exponential growth after a period of slowing due to critical telomere shortening. The explanation for this phenomenon is that some cells engage telomerase-independent pathways to maintain telomeres that allow them to become "survivors." The model takes into account random variation in individual cell cycle times and other factors and leads to estimation of parameters such as the probability of an individual cell becoming a survivor.

The model uses the paradigm of the general branching process (Haccou et al. 2005) and assumes that individuals reproduce by budding. The times between consecutive budding events (i.e., the cell cycle times) are assumed to be independent random variables with the same distribution. A parent cell produces progeny cells at different times during its life which is relevant to budding yeast and different from the usual binary fission models. An individual yeast cell contains 16 chromosomes, so it has 32 telomeres. In cells that express telomerase, the number of telomeric repeats present varies to a certain extent from end to end resulting in a distribution of telomere lengths around a strain-specific point (Shampay and Blackburn 1988). In the absence of telomerase, the rate of loss of telomeric DNA follows a probability distribution over a range from three to five base pairs per end per cell division (Lundblad and Szostak 1989; Singer and Gottschling 1994). The "telomere unit" is defined to be four base pairs and it is assumed that one telomere unit is lost per division. The type of a cell is the number of remaining telomere units.

A population starts from a single telomerase-deficient cell of type n. The cell produces one progeny cell upon completion of its cell cycle. All telomeres have shortened by one unit in the preceding round of DNA replication and are randomly allocated to the parent cell and to the progeny cell. It is reasonable to assume that both parent and progeny, after division, have type $n - 1$. When a cell reaches type 0 (the critical telomere length), it stops dividing and becomes senescent, unless an alternative mechanism of telomere maintenance is established. Senescent cells remain in the population but do not further reproduce. It is assumed that cell cycle times are independent random variables with a common cumulative distribution function (cdf) F. The main quantity of interest is the number of cells in the population at time t, which we denote by Z_t. The population starts from a single cell of type n

at time $t = 0$. In this basic model, the expected value of Z_t is

$$E[Z_t] = 1 - F(t) + \sum_{k=1}^{n-1} 2^k \left(F^{*k}(t) - F^{*(k+1)}(t)\right) + 2^n F^{*n}(t), \tag{7.55}$$

where F^{*k} denotes k-fold convolution of F, that is, the cdf of the sum of k cell cycle times. The factor 2^k is the expected number of cells in the kth generation. Each cell in the kth generation is present in the population if the sum of k cell cycle times is less than t but the sum of $k+1$ cell cycle times is greater than t. As the probability of this event is $F^{*k}(t) - F^{*(k+1)}(t)$, the expected number of cells from the kth generation that are present at t equals

$$2^k \left(F^{*k}(t) - F^{*(k+1)}(t)\right) \tag{7.56}$$

and, noting that cells of type 0 do not further reproduce, summing over k gives the expression for $E[Z_t]$. $E[Z_t] \to 2^n$ as $t \to \infty$ so 2^n is the final number of cells.

To account for the restored exponential growth, it is assumed that cells that have reached type 0 have the possibility to turn into "survivors" or become senescent. The expression for $m(k)$, the expected number of cells in the kth generation, remains the same for $k \leq n$, but for $k > n$ it changes since cells of type 0 may now escape senescence and keep reproducing. Cell of type 0 becomes a survivor with probability p and the survivor status is inherited by all of its progeny. Thus, each survivor starts a population of survivors where telomere length is generally maintained and we assume the only limiting factor is the "absolute" (not related to telomere length) replicative lifespan n_0. The nonsurvivors turn senescent.

The final expression is provided in Olofsson and Bertuch (2010) Proposition 2.2, which can be restated as follows:

Proposition *Let $M(k)$ denote the expected number of cells in generation k in a branching process where cell cycle times are independent with cdf F, the initial telomere length is n, the replicative lifespan is n_0, the probability of becoming a survivor is p. Then*

$$M(k) = \begin{cases} 2^k, & \text{for } k \leq n_0 \\ m(k), & \text{for } n_0 < k \leq n \\ d(n) + (1-p)s(n) + p\sum_{i=1}^{n_0} m_{i,k-n}, & \text{for } k > n \end{cases} \tag{7.57}$$

and the expected number of cells in the population at time t is

$$E[Z_t] = 1 - F(t) + \sum_{k=1}^{\infty} M(k)(F^{*k}(t) - F^{*(k+1)}(t)). \tag{7.58}$$

For brevity, we omit the definitions of functions $m(k)$, $s(n)$, and m_{ij} which are related to the expected number of progeny of cells of various type and refer the reader to the original paper.

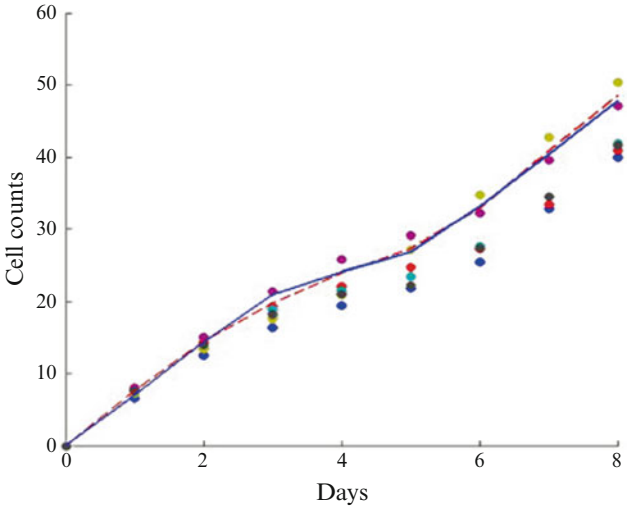


Fig. 7.6 Cell counts (logarithmic scale) of seven yeast cultures for the first eight days, fitted by model curve. (Source: Olofsson and Bertuch 2009)

We reproduce Fig. 7.6 from Olofsson and Bertuch (2010), which shows the sigmoidal shape characteristic of the population growth becoming first arrested and then restored. The fit has been achieved assuming $p = 0.001$.

7.8 Application: Structured Cell Population Models

Structured population models describe proliferation of populations taking into account distributions of variables characterizing individuals. In the context of cell populations, examples of structural variables are cell mass, levels of biochemical constituents such as RNA or proteins, degree of cell maturation or differentiation, etc. A frequently used way of modeling structured cell populations is by means of partial differential equations (PDE) of transport type. One of the most general models of this type was analyzed by Webb (1987). Another comprehensive reference is the book by Diekmann and Metz (1987). An alternative approach employs branching processes and more general stochastic processes (Arino and Kimmel 1993). Type space should be rich enough to accommodate a structure variable x, varying in a continuum, e.g., in an interval or another subset of the real line. This can be accomplished using general branching processes.

Another class of models describes the expected values of stochastic (branching) processes of cell proliferation. These models employ integral equations of renewal type, including type-transition laws in the kernel functions under the integral sign. Examples of such models are those by Tyson and Hannsgen (1984), Kimmel et al. (1984), Arino and Kimmel (1987), and by others.

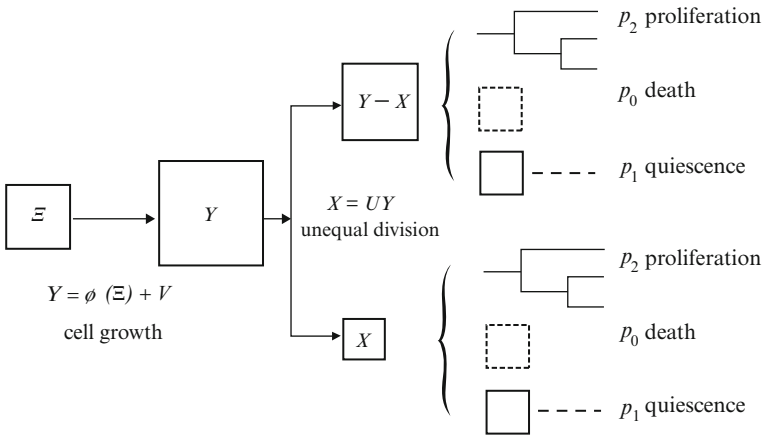


Fig. 7.7 A schematic representation of the cell cycle model. A cell of size $\tilde{X}$ at its birth grows during a single generation to size $Y = \phi(\tilde{X}) + V$, where ϕ is an increasing function and V is a nonnegative random variable with cumulative distribution G. In mitosis, the cell divides into two daughters of unequal sizes X and $Y - X$ according to the rule $X = UY$, where U is a random variable on $[0, 1]$ (independent of V) with a symmetric distribution H. Each of the daughter cells, independently, starts growing with probability p_2, dies with probability p_0, or becomes quiescent with probability p_1 ($p_0 + p_1 + p_2 = 1$). (Source: Kimmel and Axelrod 1991)

7.8.1 A Model of Unequal Division and Growth Regulation in Cell Colonies

The model introduced by Kimmel and Axelrod (1991) unifies features of a time discrete GW branching process and those of a deterministic model of cell cycle regulation introduced by Kimmel et al. (1984). It is a generalization of the example in Sect. 3.2. A schematic representation is depicted in Fig. 7.7.

Cell Growth. A cell of size (mass, volume) X_0 at its birth grows during a single generation to size $Y = \phi(X_0) + V$, where V is a nonnegative random variable with given cumulative distribution G. Function ϕ represents the size regulation mechanism; it is assumed nondecreasing which means that progeny cells larger at birth are also larger at division. However, specific assumptions on ϕ ensure that any deviation from the average size, if present at the birth of the cell, decreases during cell growth. For mathematical simplicity, it is assumed that proliferating cells have identical lifetimes and that the lifetimes of the quiescent cells are infinite.

Unequal Division. In mitosis the parent cell of size Y divides into two progeny of unequal sizes X and $Y - X$. It is assumed that the size of one of the progeny cells is equal to $X = UY$, where U is a random variable with values in $[0, 1]$, independent of Y and V, with a symmetric distribution H. Formally,

$$\mathrm{P}\{U \le u\} = H(u) = 1 - H(1 - u), \quad H(0) = 0, \quad H(1) = 1.$$

The size of the other progeny is $(1 - U)Y$.

Proliferation. Each of the progeny cells chooses its own pathway, independent of its parent's size, of its own size and of the pathway chosen by the other progeny, based on a purely random rule. With probability p_2, the cell starts growing and initiates a pedigree, with probability p_0 it dies, or with probability p_1 it becomes quiescent, i.e., continues to exist without either growing or dying.

Independence. Due to the assumed independence of cell death and quiescence from growth regulation and unequal division, one prediction of the present model is that the distribution of number of cells per colony does not depend on the birth size of initial cell. In particular this implies independence between number of cells within a colony and birth sizes of cells within the colony, at any time after the initiation. This is consistent with experimental observations (see Fig. 7.8 and Kimmel and Axelrod 1991).

Let us note that because of independence between cell proliferation, quiescence and death on one hand, and cell growth and unequal division on the other, the total count of proliferating and quiescent cells obey the laws of the GW process in the example in Sect. 3.2. Therefore, we focus here on the size structure of the process.

Let

$$M_i(x, x_0) = \mathrm{E}[N_i(x, x_0)]$$
$$R_i(x, x_0) = \mathrm{E}[Q_i(x, x_0)]$$

denote the expected numbers of proliferating and quiescent cells with birth-sizes not exceeding x, in the ith generation of the process started by a single cell with birth-size x_0. These counting functions describe the cell size structure of the population.

Theorem 7.4 *Under suitable hypotheses* (Kimmel and Axelrod 1991)*, the following recurrences are satisfied:*

$$M_{i+1}(x, x_0) = 2p_2 \int_0^\infty \int_x^\infty H(\frac{x}{y}) d_y M_i[\phi^{-1}(y - v), x_0] dG(v),$$
$$M_0(x, x_0) = 1(x - x_0). \tag{7.59}$$

and

$$R_i(x, x_0) = \frac{p_1}{p_2} \sum_{j=1}^{i} M_j(x, x_0),$$
$$R_0(x, x_0) = 0. \tag{7.60}$$

Dynamics of Cell Size Distributions The experimental data available for comparison with the model are the empirical distributions of cell size (understood as volume) in the same experimental system as presented in the chapter concerning the GW process. Figure 7.9 depicts the cumulative distribution of sizes of the measured NIH progeny cells. The corresponding distribution of the NIH(*ras*) cells is indistinguishable.

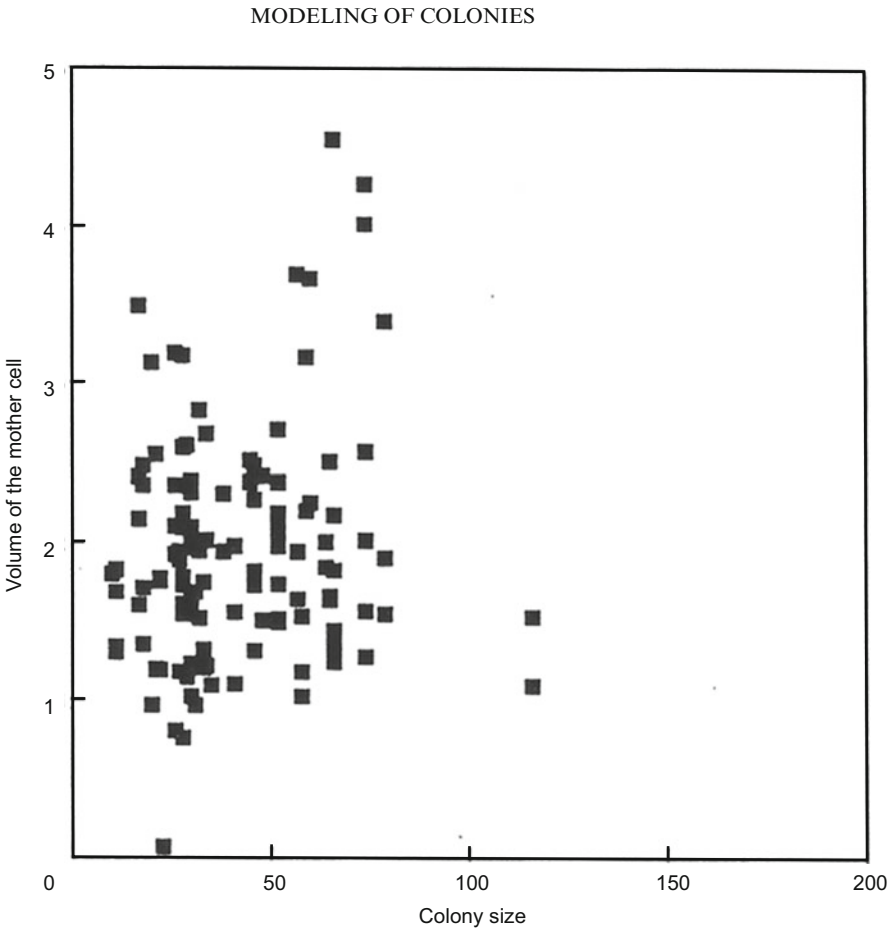


Fig. 7.8 Size at division of cells in colonies with different number of cells. Colonies were grown for 4 days from single cells. Then, for each colony, cells were counted and the sizes of pairs of cells after division were determined. Up to three pairs of dividing cells per colony were recorded. The sum of the volumes of daughter cells is given as the volume of the mother cell. No dependence of cell size at division on colony size is apparent. (Source: Kimmel and Axelrod 1991)

The question to answer by mathematical modeling is the following: Using the empirical distribution H of inequality of cell division and a mathematical form of the growth function ϕ, is it possible to reproduce the observed size distribution at the end of the experiment?

Figure 7.9 depicts the evolution of distributions of cell size modeled using Eqs. (7.59) and (7.60). To obtain cumulative distributions, the counting functions have been normed. If it is assumed that the size of the founder cell of the colony was

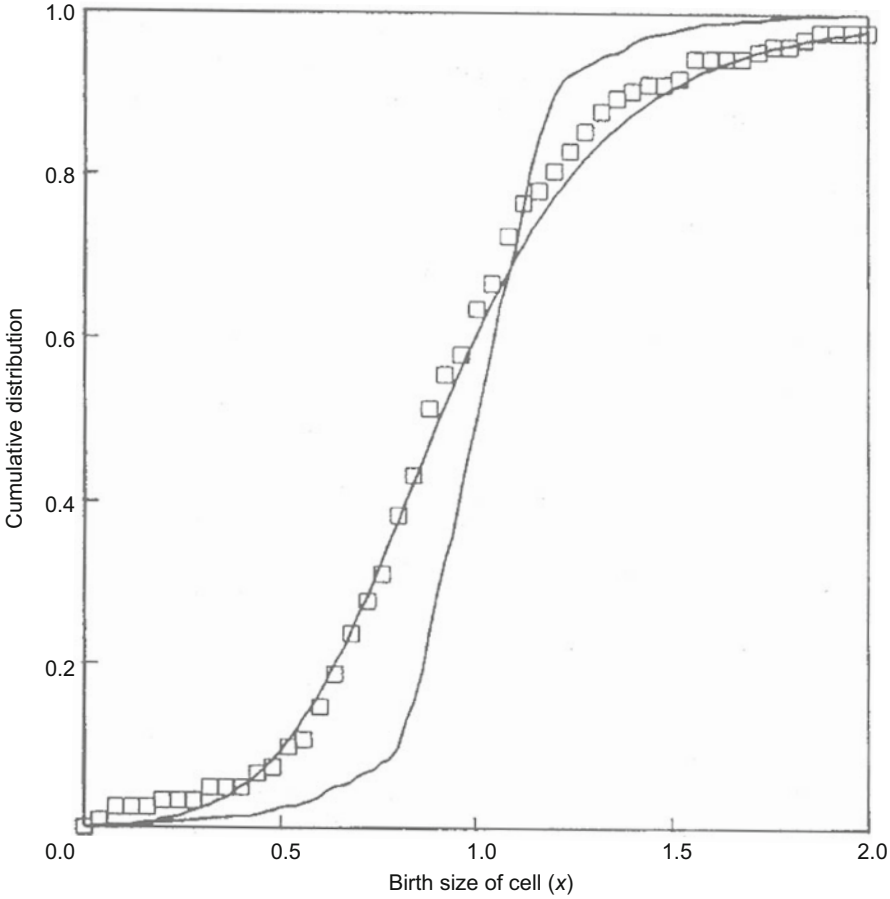


Fig. 7.9 Distributions of cell sizes for NIH cells. Cell sizes were microscopically determined immediately after division. Symbols represent the observed cumulative frequency of the daughter cell sizes. Continuous lines are expected size distributions generated by the model, after one and eight generations, starting with a single founder cell of size 1 unit. The modeled distribution of cell sizes after eight generations closely resembles the empirically observed distribution. (Source: Kimmel and Axelrod 1991)

$x_0 = 1$, the modeled colony size distribution at generation eight (end of experiment) is close both to the limit distribution and to the empirical distribution. If the founder cell size is assumed very small or very large, the convergence to the limit distribution is still satisfactory within ten generations (not shown). The fit provided by the model is satisfactory for cell size distributions as well as colony size distributions (Sect. 3.2).

7.8.2 Cell Cycle Model with Cell-Size Control, Unequal Division of Cells, and Two Cell Types

Among many laboratory and mathematical models of structured cell populations, the system introduced by Sennerstam (1988) stands out because it encompasses almost all features encountered in such systems. Also, it gave rise to a series of studies ranging from laboratory investigation, through theorizing and computer modeling, to mathematically advanced models in the form of renewal equations and general branching processes.

Similarly as for the model in the preceding section, Sennerstam's (1988) studies were motivated by the observation made in the 1960s that the partition of mass to daughter cells at mitosis is asymmetric. Furthermore, it was suggested that such unequal distribution of metabolic constituents at mitosis contributes to the dispersion of cell generation times and cell masses in a population. Various theories were put forward concerning the mechanisms of regulation of generation time and cell growth rate, given cell's birthmass and other factors (among them Darzynkiewicz et al. 1979, 1982; Cooper 1984; Kimmel et al. 1984).

In Sennerstam (1988), cultured PCC3 embryonal carcinoma (EC) cells were studied in order to evaluate their protein content. There exists a considerable intraclonal intermitotic time heterogeneity found in undifferentiated PCC3 EC cells. It was concluded that the postmitotic difference in mass (protein) between sister cell pairs has an influence upon variation in cell-cycle time duration when comparing sister cell pairs. This offered an explanation for the randomly distributed difference repeatedly found between sister cell generation times. In spite of this, there was no correlation seen between the mass difference found between sister cell pairs postmitotically and the mass of the mother cell.

In a subsequent paper, Sennerstam and Stromberg (1988) reported the discovery of an intraclonal bimodal-like cell cycle time variation within the multipotent EC PCC3 N/1 line. The variability was found to be localized in the G1 period. Furthermore, an inverse relation between cell mass and cell generation time was found in the cell system analyzed. It was suggested that the bimodal intraclonal time variability previously reported was attributable to an intraclonal shift between two types of cell-growth-rate cycles.

To explain the findings, Sennerstam and Stromberg (1995, 1996) used the so-called continuum model (Cooper 1984). The model is based on the idea that DNA replication and cell growth are two loosely coupled subcycles. After division (generally asymmetric), a cell proceeds through the G1 phase until it reaches a checkpoint characterized by a threshold mass. At this moment, the DNA synthesis is triggered and the time to division and further mass increase (or a growth rate) are determined. The growth continues after division at the same rate, etc. In this way, the division cycle (from one cell division to another) is only partly coupled with the growth cycle, since adjustments to the growth rate are made only at the $G1/S$ boundary checkpoints. Thresholds and rates have stochastic components and consequently the mass-at-division regulation is not perfect.

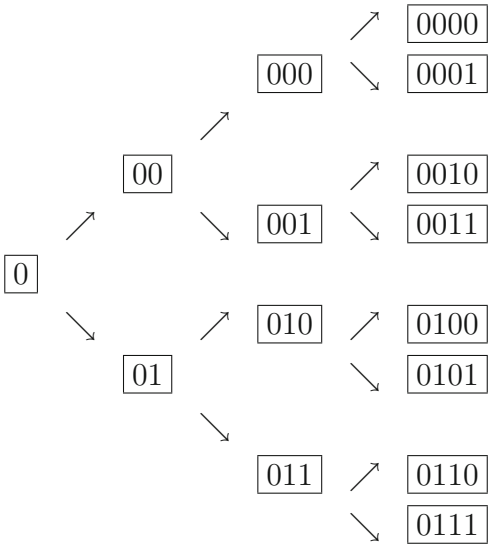
Sennerstam's measurements described above were used by Kimmel and Arino (1991) to build a mathematical model, equivalent to expected-values equations for a branching process. An extremely simplified version was already mentioned in Sect. 6.3. The model takes into account *cell size regulation* (cells grow between divisions, at certain mass they decide to divide), *unequal division* (some cell constituents do not split equally between progeny cells), and *differentiation* (cells switch off/on some of their genes to specialize in a required direction).

The following detailed observations were listed in Sennerstam (1988) describing growth characteristics of immortalized embryonic cells:

- Using mitotic detachment technique, it was established that the coefficient of variation of the mass of progeny cells exceeded the coefficient of variation of parent cells by about 4 %, i.e.,

$$cv_{\text{progeny mass}}/cv_{\text{parent mass}} \cong 1.04.$$

- Using time-lapse measurements, the distributions of the generation times of related cells were determined. The indexation of generation (interdivision) times and other variables describing cell pedigrees is explained by the following diagram:



  - α-curve $= f_{T_0}(\tau)$, the distribution of cell lifelengths, was found to be *bimodal*.
  - β_1-curve $= f_{|T_{00}-T_{01}|}(\tau)$, the distribution of differences of sib cells lifelengths, was found to be *unimodal*.
  - β_2-curve $= f_{|T_{000}-T_{011}|}(\tau)$, the distribution of differences of the first cousin cells lifelengths, was found to be bimodal.
  - β_3-curve $= f_{|T_{0000}-T_{0111}|}(\tau)$, the distribution of differences of the second cousin cells lifelengths, was found to be unimodal.

- Furthermore, correlation coefficients between generation times of related cells were estimated:
  - Parent–progeny, $\rho_{T_0,T_{00}} = 0.77$
  - Sib–sib, $\rho_{T_{00},T_{01}} = 0.95$
  - Cousin–cousin, $\rho_{T_{000},T_{011}} = 0.41$

Kimmel and Arino (1991) proposed the following model, also based on Cooper's (1984) continuum hypothesis, to explain the observations:

- There exist two types of cells: $\begin{cases} \text{type } \boxed{1}, & \text{smaller and faster cycling,} \\ \text{type } \boxed{2}, & \text{larger and slower cycling.} \end{cases}$

- Growth of cell mass between divisions proceeds at a constant rate r. Specifically,
  - Cell of type i and initial mass y, grows in the $G1$ phase to a random threshold size

$$w_i \sim h_i(\,\cdot\,), \quad w_1 \overset{(d)}{<} w_2,$$

    where the stochastic inequality between $w_1$and w_2 is equivalent to the same relation between their tail distributions, i.e., $\mathrm{P}[w_1 > x] < \mathrm{P}[w_2 > x]$.
  - Then it continues through phases $S + G2 + M$ for a constant time τ, i.e., the total duration of the cell cycle is equal to

$$T = \frac{w_i - y}{r} + \tau,$$

  - And grows to the predivision mass x

$$x = w_i + r\tau.$$

- Switching between types: At the checkpoint on the G1/S boundary, it is decided if the type of progeny (both) is the same as the parent, or not

$$\Pr[i \to j] = p_{ij}.$$

This is the "supramitotic regulation," i.e., decisions are made at a checkpoint inside the division cycle.

- Unequal division: Parental cell of mass x_0 divides into progeny of masses y_{00} and y_{01}. Asymmetry of division can be represented by multiplication of x_0 by a random variable u_0, with distribution with support in $[0, 1]$, as represented in the diagram below,

The model stated above explains the observations of Sennerstam (1988). Let us assume that the cell population is in the state of asynchronous exponential growth, i.e.,

$$\begin{pmatrix} N_1(t) \\ N_2(t) \end{pmatrix} = C \begin{pmatrix} \widetilde{p}_1 \\ \widetilde{p}_2 \end{pmatrix} \exp(\lambda t),$$

$$\boxed{\mathrm{i}|\mathrm{j}}_0 \begin{cases} \nearrow \boxed{\mathrm{j}|\cdots}_{00} & y_{00} = u_0 x_0 \\ \searrow \boxed{\mathrm{j}|\cdots}_{01} & y_{01} = (1-u_0)x_0 \end{cases},$$

$$u_0 = f_0(u)^{\mathsf{c}} \quad f_0\,(u) = f_0(1-u).$$

where $N_i(t)$ is the number of cells of type i at time t. To address the bimodality of the α-curves, let us suppose that a cell of type i is the progeny of a cell of type j. This occurs with probability

$$\tilde{p}_j p_{ji}.$$

Then

$$T_0|i \sim \tau + \frac{w_i - y_j}{r} = \tau + \frac{w_i - (w_j + r\tau)u}{r}.$$

This latter, under equal division, $u = 1/2$, reduces to

$$\frac{\tau}{2} + w_i - \frac{w_j}{2}.$$

Now, let us assume that the distribution of u is tightly concentrated around $1/2$. If in addition, $\tilde{p}_1 \cong \tilde{p}_2 \cong 1/2$, and p_{12} and p_{21} are small, then the two dominating modes of the distribution of random variable T_0 are approximately located at

$$\frac{\tau}{2} + \frac{w_i}{2},\ i = 1, 2.$$

Unimodality of distributions of differences of lifelengths of sib cells (the β_1-curves) follows since

$$T_{00} - T_{01} = \left[\tau + \frac{w_{00} + (w_0 + r\tau)u_0}{r}\right] - \left[\tau + \frac{w_{01} + (w_0 + r\tau)(1-u_0)}{r}\right],$$

which, under $u = 1/2$, is equal to

$$\frac{w_{00} - w_{01}}{r},$$

so that $|T_{00} - T_{01}|$ has the only mode at zero.

Bimodality of first cousin lifelength difference distributions (the β_2-curves) follows since

$$T_{000} - T_{010} = \frac{1}{r}\left[(w_{000} - w_{010}) - r\tau(u_{00} - u_{01}) - (w_{00}u_{00} - w_{01}u_{01})\right].$$

Again, under $u = 1/2$, this is equal to

$$\frac{1}{r}\left[\underbrace{(w_{000} - w_{010})}_{w_i - w_j} - \frac{1}{2}\underbrace{(w_{00} - w_{01})}_{\text{same distribution}}\right],$$

so $|T_{000} - T_{010}|$ has one mode at 0 and another (smaller) at $|w_1 - w_2|$.

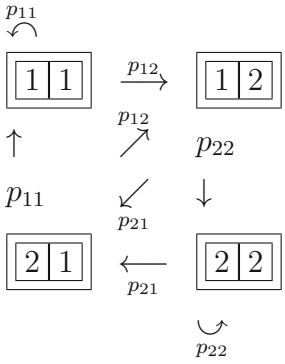
In addition to the above, Kimmel and Arino (1991) carried out computations of the correlations of related cells, under assumption $p_{11} = p_{22}$, with all other parameters fitted to data. The conclusions are as follows:

- Parent–progeny correlation is negative if p_{11} small (frequent switching of type), positive if p_{11} large.
- Sib–sib correlation is more or less stable (same w).
- Cousin–cousin correlation large if p_{11} large (type rarely changed) and if p_{11} small (type likely to be the same each second generation).

An important theoretical problem concerns the dynamics of cell proliferation in this case: How to reconcile the division cycle with the growth cycle and with unequal division and the random decisions to switch cell type (these latter assumed to be taken at the $G1/S$ checkpoint)? It seems convenient to introduce four types of cells, indexed by pairs $(i, j)_{i,j=1,2}$

$$\boxed{\boxed{\text{i}}\boxed{\text{j}}} = \text{type } i \text{ that decided on progeny type } j.$$

Transitions reduced to "decision taken at birth" are depicted in the following diagram:



Using these transitions allows writing straightforward balance equations for expected densities of flow rates between types. Suppose

$$n_{ij}(t, y)dtdy$$

is the expected flow of (i, j) progeny of size in $(y, y + dy)$ into growth phase, in time interval $(t, t + dt)$, then

$$n(t, y) = 2r \int f(x, y)H(x - r\tau) \int n[t - (\tau + \sigma), x - r(\tau + \sigma)]d\sigma dx$$

where

$$n(t,y) = \begin{pmatrix} n_{11}(t,y) \\ n_{12}(t,y) \\ n_{21}(t,y) \\ n_{22}(t,y) \end{pmatrix}$$

$$H(w) = \begin{pmatrix} p_{11}h_1(w) & 0 & p_{11}h_2(w) & 0 \\ p_{12}h_1(w) & 0 & p_{12}h_2(w) & 0 \\ 0 & p_{21}h_1(w) & 0 & p_{21}h_2(w) \\ 0 & p_{22}h_1(w) & 0 & p_{22}h_2(w) \end{pmatrix}$$

As demonstrated in Kimmel and Arino (1991), this evolution equation generates a semigroup of operators on $X = L^1(E)$, the space of functions integrable on E, i.e., such that $n(t) = \int_{y \in E} n(t,y)\mathrm{d}y < \infty$, where the set E of admissible sizes of progeny is defined by specific assumptions on distributions $h_1(w)$ and $h_2(w)$,

$$G(t) : X \ni n_0 \longrightarrow n_t.$$

The main mathematical problem is to show that asynchronous exponential growth exists. It is sufficient to show that spectrum of $G(t)$ has a dominating eigenvalue $\exp(\lambda t)$ and that the corresponding generalized eigenspace is one-dimensional. This is true since $G(t)$ is eventually compact. Projection of solution n_t on the generalized eigenspace dominates all other solutions and yields

$$n(t) \sim \exp(\lambda t)$$

and

$$N(t) \sim \exp(\lambda t)$$

as desired.

Alexandersson (1999) proposed a largely equivalent description in the form of a general branching process (Sect. C.1). The process has type space

$$\{11, 12, 21, 22\} \times \{R_+\}$$

with reproduction measure determined by the transition rules above. Finding Malthusian parameter for the process is equivalent to solving the characteristic equation for the dominant eigenvalue in the model by Kimmel and Arino (1991). Then the problem is to demonstrate conservativeness of the reproduction measure. This is in some sense equivalent to proving results concerning the eigenspace corresponding to the dominant eigenvalue of the semigroup $G(t)$.

Alexandersson (1999) considers various versions of the regulation mechanism of the cell growth, some of them involving variable growth rates. Those who venture to read both Kimmel and Arino (1991) and Alexandersson (1999), will see that the general branching process methodology makes the modeling process conceptually more straightforward.

Another example of cell cycle analysis is the paper by Guo et al. (2013), who employ a branching-process-based algorithm to deconvolve transcription data in synchronized cell experiments in yeast. Cells released from a synchrony block proceed through the cell cycle and then divide, potentially multiple times, loosing synchrony in the process. The authors introduce a two-type Bellman–Harris process, in which the types are (i) mothers and (ii) daughters. This distinction is due to the fact that the $G1$ phase of the daughters is structured differently from that of the mothers. By fitting the model on the transcription data in different cell cycle phase, it is possible effective deconvolve transcription from cell cycle duration differences, and obtain more accurate estimates of the timing of transcription events.

7.9 Application: Yule's Evolutionary Process

We paraphrase Yule's branching process model of evolution, as cited in Harris (1963). The model concerns the evolution of two basic taxonomic units, species and genera, within a single family. The following assumptions define the model:

1. Two types of objects are considered.
   (a) *Species*: This is the smallest taxonomic unit. Different species are reproducively separated, i.e., if individuals belonging to different species are crossed they do not produce fertile progeny.
   (b) *Genus*: Species are grouped in genera. The biological distance separating different genera is larger than that separating different species.
2. Within a single genus, the collection of species evolves as an age-dependent branching process with exponential lifetime distributions with parameter λ and the pgf of the number of progeny equal to $h(s) = s^2$ (i.e., each speciation event produces exactly two new species).
3. The collection of genera evolves as an age-dependent branching process with exponential lifetime distributions with parameter γ. However, at each ramification a new genus evolves which has exactly one species and the old genus continues unchanged. This asymmetry is caused by the fact that a new genus arises from a major evolutionary rearrangement within a single species of the old genus.

The object defined is a sort of a "branching process within a branching process" (Fig. 7.10)

Let $N(t)$ denote the number of genera existing at time t, and $M_i(t)$ the number of species existing at t in the ith genus. Then, the process can be defined as the vector

$$\mathcal{Z}(t) = \{M_1(t), M_2(t), \ldots, M_{N(t)}(t)\}, \quad t \geq 0.$$

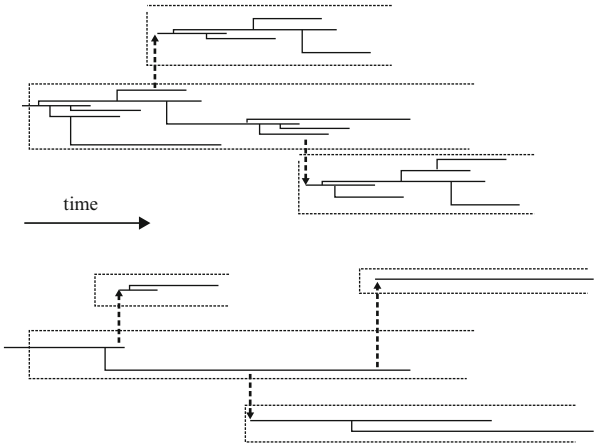


Fig. 7.10 Two sample paths of the Yule's process, low and high value of the $\frac{\gamma}{\lambda}$ ratio, respectively. Branching of species is represented by *continuous lines*. Boundaries of genera are represented by *dotted line* "tubes," and branching of genera by *arrows*. (Source: Kimmel and Mathaes 2010b)

Finding a comprehensive description for $\mathcal{Z}(t)$ is quite complicated, since it is an age-dependent branching process with infinitely many particle types. However, we are interested in answering a very particular question regarding the process: What is the rate of evolution of new genera compared to evolution of new species, as measured by the ratio of γ to λ?

Let us notice (see Fig. 7.10) that high $\frac{\gamma}{\lambda}$ ratio yields a relatively high frequency of *monotypes,* i.e., genera with one species. Therefore, by examining the population-based proportion of monotypes at given time t, it seems possible to estimate the $\frac{\gamma}{\lambda}$ ratio. This gives a chance to infer about the dynamics of the evolutionary process without actually observing it. It is particularly important in view of patchiness and discontinuity of paleontological evidence.

We develop a model with only two classes of genera, *class 1* genera containing monotypes and *class 2* genera with more than one species. The number, at time t, of genera of class i is denoted by $Z_i(t)$, $t = 1, 2$. The joint pgf of the pair $(Z_1(t), Z_2(t))$ in the process started by a single class i genus is denoted by $F_i(s_1, s_2; t)$. We show the process is a two-type *Markov* age-dependent process.

The lifetime distribution of a class 2 genus is $G_2(\tau) = 1 - \mathrm{e}^{-\gamma\tau}$. The joint pgf of the class 1 and 2 progeny of such genus is $h_2(s_1, s_2) = s_1 s_2$.

A class 1 (monotype) genus transforms into a class 2 genus after a time τ' distributed exponentially with parameter λ, because of a speciation event; independently, it splits into two genera after a time τ'' distributed exponentially with parameter γ. The minimum of these two times is τ distributed exponentially with parameter $\lambda + \gamma$. If $\tau' < \tau''$, which happens with probability $\frac{\lambda}{\lambda+\gamma}$, then the "progeny" pgf is s_2. Otherwise, it is s_1^2. Eventually, $G_1(\tau) = 1 - \mathrm{e}^{-(\lambda+\gamma)\tau}$ and $h_1(s_1, s_2) = \frac{\lambda}{\lambda+\gamma} s_2 + \frac{\gamma}{\lambda+\gamma} s_1^2$. For a two-type age-dependent branching process, the

pgf equations are a straightforward generalization of Eq. 4.6:

$$\frac{\partial F_i(s;t)}{\partial t} = \lambda_i\{h_i[F_1(s,t), F_2(s,t)] - F_i(s,t)\}; \quad t \geq 0,\ F_i(s;0) = s_i;\ i = 1,2. \tag{7.61}$$

This yields, in our case,

$$\frac{\partial F_1}{\partial t} = \gamma F_1^2 + \lambda F_2 - (\lambda + \gamma)F_1, \tag{7.62}$$

$$\frac{\partial F_2}{\partial t} = \gamma F_1 F_2 - \gamma F_2. \tag{7.63}$$

Using methods similar as in the application to clonal resistance theory and in particular a variant of Theorem 4.2, an explicit solution of the above system is obtained:

$$F_2 = \frac{s_2(\lambda+\gamma)\mathrm{e}^{-\gamma t}}{[\lambda(1-s_2)+\gamma(1-s_1)] + s_2(\lambda+\gamma)\mathrm{e}^{-\gamma t} + (s_1-s_2)\gamma\mathrm{e}^{-(\gamma+\lambda)t}},$$

$$F_1 = (1 + \frac{s_1 - s_2}{s_2}\mathrm{e}^{-\lambda t})F_2.$$

Suppose the process (i.e., given family of genera) was started at time 0 by a monotypic genus. Then the expected numbers of monotypic and polytypic genera at time t are given by

$$M_{11}(t) = \frac{\partial F_1(s;t)}{\partial s_1}_{|s=(1,1)} = \frac{\gamma}{\lambda+\gamma}\mathrm{e}^{\gamma t} + \frac{\lambda}{\lambda+\gamma}\mathrm{e}^{-\lambda t}, \tag{7.64}$$

$$M_{12}(t) = \frac{\partial F_1(s;t)}{\partial s_2}_{|s=(1,1)} = \frac{\lambda}{\lambda+\gamma}(\mathrm{e}^{\gamma t} - \mathrm{e}^{-\lambda t}). \tag{7.65}$$

Eventually, the expected proportion of monotypic genera in the family which is old enough, is equal to

$$p = \lim_{t\to\infty} \frac{M_{11}(t)}{M_{11}(t) + M_{12}(t)} = \frac{\gamma}{\lambda+\gamma}. \tag{7.66}$$

In the book of Harris (1963), an example is quoted of a family of beetles with 627 genera comprising 9997 species, $p = 34.29\%$ of the genera being monotypes. From this, it can be estimated that $\lambda/\gamma = 1.9$.

7.10 Application: Branching Process with Infinite-Allele Mutations

7.10.1 Proliferation of Alu Repeats

About 11 % of the human genome consists of more than one million repeated sequences of about 300 base pairs, called *Alu* elements. *Alu*'s are retrotransposons;

they can produce inexact mutant copies with different base sequences at different locations. Each *Alu* element with a different sequence can be considered a different allele. Comparison of the base sequences of many *Alu* elements suggests that there are different families of similar sequence each derived from a different parental sequence. The observation that progeny sequences within some families are more similar to each other than the progeny sequences within other families suggests that some families are younger than others and that the number of families in the human genome is still growing.

Discrete Branching Process of Griffiths and Pakes with Infinite Allele Mutations

The proliferation and mutation of *Alu* sequences can be described mathematically as a branching process. The growing number of *Alu* families in the human genome can be accounted for by a supercritical branching process. The large number of different progeny sequences (alleles) within each family can be accounted for by a discrete-time branching process with an infinite number of alleles. Griffiths and Pakes (1988) modified the standard BGW branching process to allow infinitely many different alleles, and derived a limit result for the expected proportion of each allele within a family.

In the application discussed here, the different alleles are identified by different base sequences, with differences that may be as small as a single base. From time $t = 0$, the clone of progeny sequences evolves in time according to a single-type branching process. With probability μ per time step, a sequence mutates and initiates a clone of a new, previously nonexistent, sequence which in turn evolves according to the same rules. As a result, a set of clones of different alleles emerges, spawning further clones, some of which may die out. We are interested in deriving, using Griffiths–Pakes (1988) theory, the expected frequencies j of each class of alleles that exist in k copies. The model prediction can be compared to observed data.

The number of individuals at $t = 0$ is defined as $Z_0 = i$. Let G_n be the collection of individuals in generation n and let Z_n denote their number. Each generation size depends on the previous generation size through the branching property

$$Z_{n+1} = \sum_{j=1}^{Z_n} \zeta_{j,n}, \tag{7.67}$$

where $\zeta_{j,n}$ are iid integer-valued random variables, which represent the number of offspring born to the jth member of G_n. The distribution of $\zeta_{j,n}$ is characterized by its pgf

$$f(s) = \sum_{k=0}^{\infty} p_k s^k, \tag{7.68}$$

where $p_k = P[\zeta_{j,n} = k]$, and it is assumed that $p_0 + p_1 < 1$, i.e., the branching process is nontrivial. We have $m = f'(1)$.

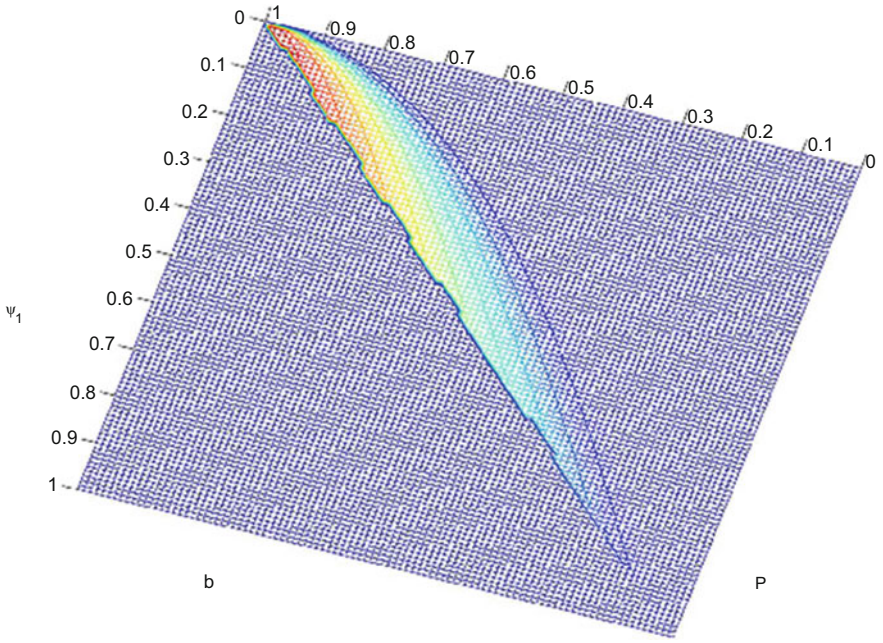


Fig. 7.11 Contour plot illustrating the influence of parameters b and p on Ψ_1, based on Griffiths–Pakes process with LF distribution. *Red*: large Ψ_1; *blue*: small Ψ_1. Range of Ψ_1-values, from 0 to 1. (Source: Kimmel and Mathaes 2010b)

If an individual produces j offspring then the number of progeny having the parental allele is distributed binomially with parameters j and $1-\mu$, hence its pgf is equal to $(\mu+(1-\mu)s)^j$. This implies that any new allele is followed by a branching process of its like-type descendants with offspring pgf $H(s) = f(\mu+(1-\mu)s)$. This process is supercritical if its expected progeny count $M = m(1-\mu)$ is greater than 1. Within this framework let us define the symbol $q_{1j}^{(r)} = \frac{(j!)^{-1}d^j H^{(r)}(s)}{ds^j}|_{s=0}$, where $H^{(r)}(s)$ is the rth iterate of pgf $H(s)$, to be equal to the probability that there are j individuals at time r in a nonmutant clone started at time 0 by a single individual. Let us denote Ψ_j the long-term expected proportion of alleles with frequency $j \geq 1$, which is the formula that we will use to compute the theoretical distribution of *Alu* allele classes for given offspring pgfs. Asymptotically, these proportions assume the form based on Griffiths and Pakes (1988), with a detailed derivation in Kimmel and Mathaes (2010a):

$$I_j = \frac{\mu \sum_{r=0}^{\infty} m^{-r} q_{1j}^{(r)}}{\mu \sum_{n=0}^{\infty} m^{-n} \left(1 - q_{10}^{(n)}\right)}. \tag{7.69}$$

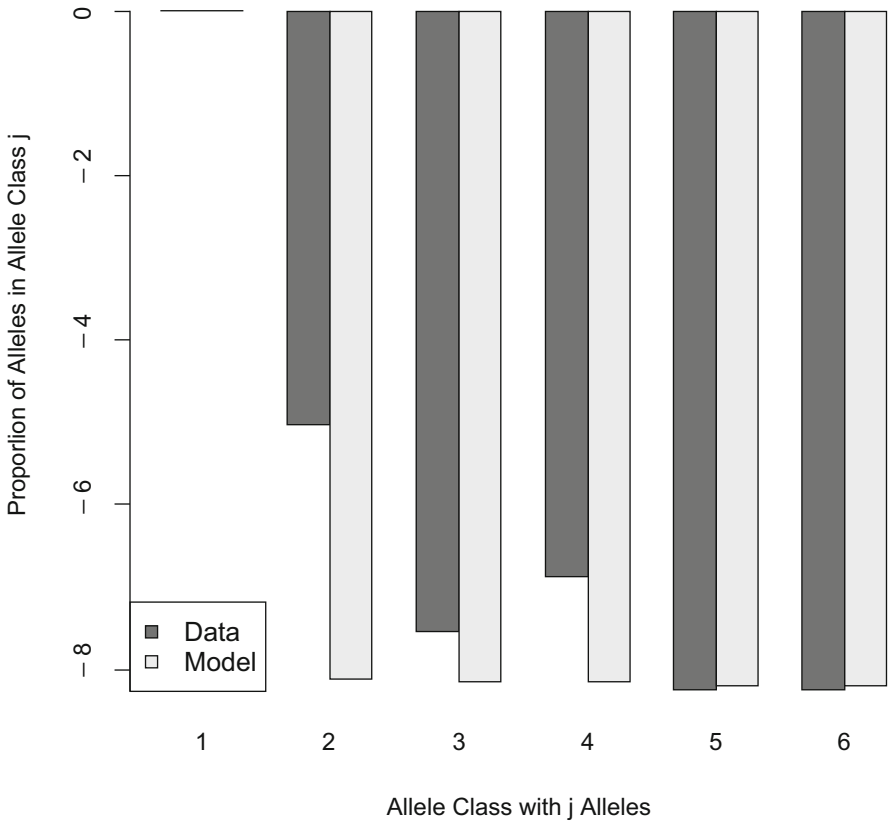


Fig. 7.12 AluYa1 data-based class frequencies against the theoretical $\{\Psi_k\}$ in log scale. Fitted by Griffiths–Pakes process with LF distribution, with $b = 0.016$, $p = 0.983$. (Source: Kimmel and Mathaes 2010b)

Linear-Fractional Offspring Distribution

The process of creation of new *Alu* repeats by retrotransposition can be naturally described by the age-dependent Markov branching process $\{Z_t\}$ (i.e., process with exponentially distributed individuals' lifelengths) with binary fission, which leads to a quadratic pgf of progeny number per individual. The rationale is that any existing *Alu* ("individual") from an active family produces two progeny (i.e., itself and a replica) at a random time moment, where "random" means that the intervals between successive fission events are independent, identically distributed random variables. Moreover, the copy may fail to reinsert into the genome. Therefore, the form of the progeny count pgf will be $\alpha s^2 + (1 - \alpha)s$, where α is the probability of successful reinsertion. If such a process is sampled at constant time intervals, the resulting discrete-time process $\{Z_k \Delta t\}$ is a GW branching process with LF pgf (Kimmel and Axelrod 2001, expression (4.14), also c.f. Arthreya and Ney 2004). A unique property

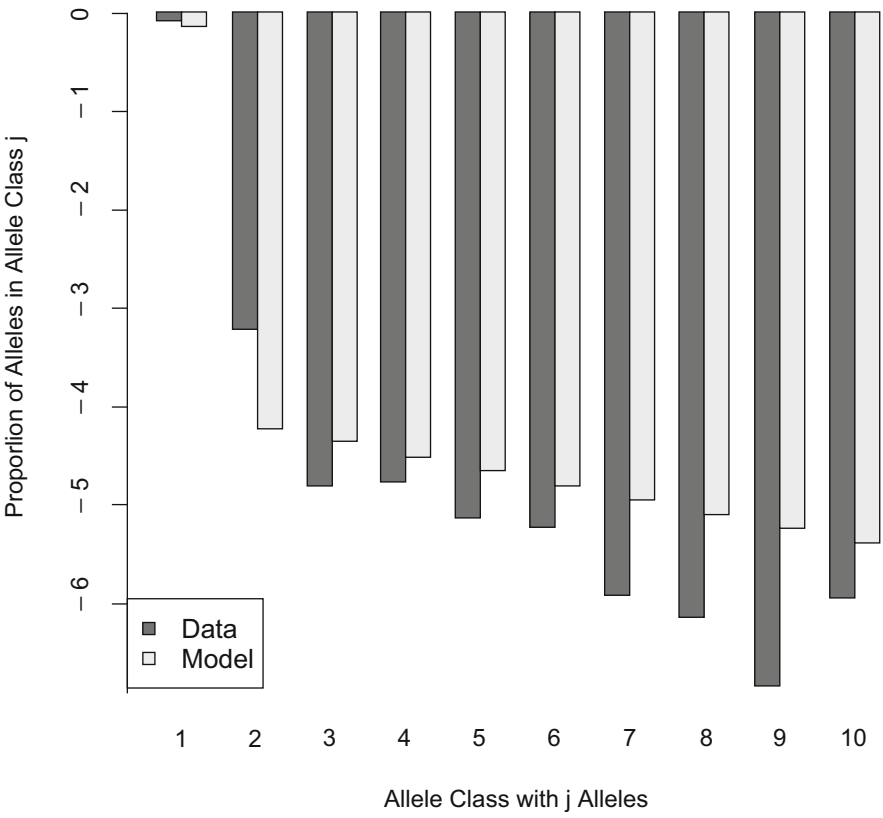


Fig. 7.13 AluYa5 data-based class frequencies against the theoretical $\{\Psi_k\}$ in log scale. Fitted by Griffiths–Pakes process with LF distribution, with $b = 0.0139$, $p = 0.861$. (Source: Kimmel and Mathaes 2010b)

of the LF case of the GW process, excluding the trivial case $f(s) = ps + q$, is that the iterations of the pgf can be computed explicitly and also are of LF form. Let us start with the offspring pgf in the LF case:

$$f(s) = 1 - \frac{b}{1-p} + \frac{bs}{1-ps}. \tag{7.70}$$

The probability distribution corresponding to this generating function is:

$$\begin{aligned} p_0 &= 1 - \textstyle\sum_{i=1}^{\infty} p_i = \frac{1-b-p}{1-p} \\ p_k &= bp^{k-1} \qquad\qquad k = 1, 2, \ldots \end{aligned} \tag{7.71}$$

The parameters b and p are subject to certain restrictions,

$$\begin{aligned} b, p &> 0, \\ b + p &\leq 1. \end{aligned} \tag{7.72}$$

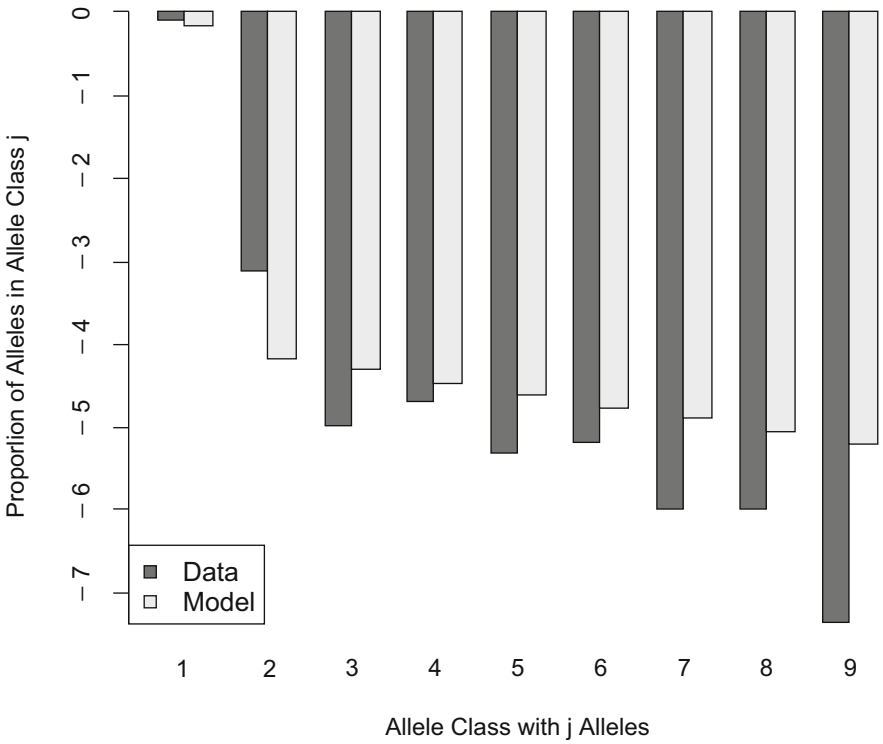


Fig. 7.14 AluYa8 data-based class frequencies against the theoretical $\{\Psi_k\}$ in log scale. Fitted by Griffiths–Pakes process with LF distribution, with $b = 0.143$, $p = 0.856$. (Source: Kimmel and Mathaes 2010b)

To ensure that this process is supercritical, i.e., $m > 1$, additional constraints on b and p are needed. The mean of $f(s)$ is

$$m = (\mathrm{d}f/\mathrm{d}s)_{|s\uparrow 1} = \frac{b}{(1-p)^2}, \tag{7.73}$$

so supercriticality yields an additional restriction on parameters b and p, $b > (1-p)^2$, or equivalently

$$p > 1 - \sqrt{b}. \tag{7.74}$$

To be more precise, we should satisfy condition $m(1-\mu) > 1$, but with μ very close to 0, the distinction is not important. As demonstrated in Kimmel and Mathaes (2010a), for the LF case we obtain the following computable expression:

$$\Psi_j = \frac{\sum_{r=0}^{\infty}(1-s_0)\frac{(m_r-1)^{j-1}}{(m_r-s_0)^{j+1}}}{\sum_{r=0}^{\infty}\frac{1}{m^r-s_0}}. \tag{7.75}$$

The infinite sums in the numerator and denominator are numerically computed. A program was written in R-language to compute the Ψ_j. Since *Alu* sequence data

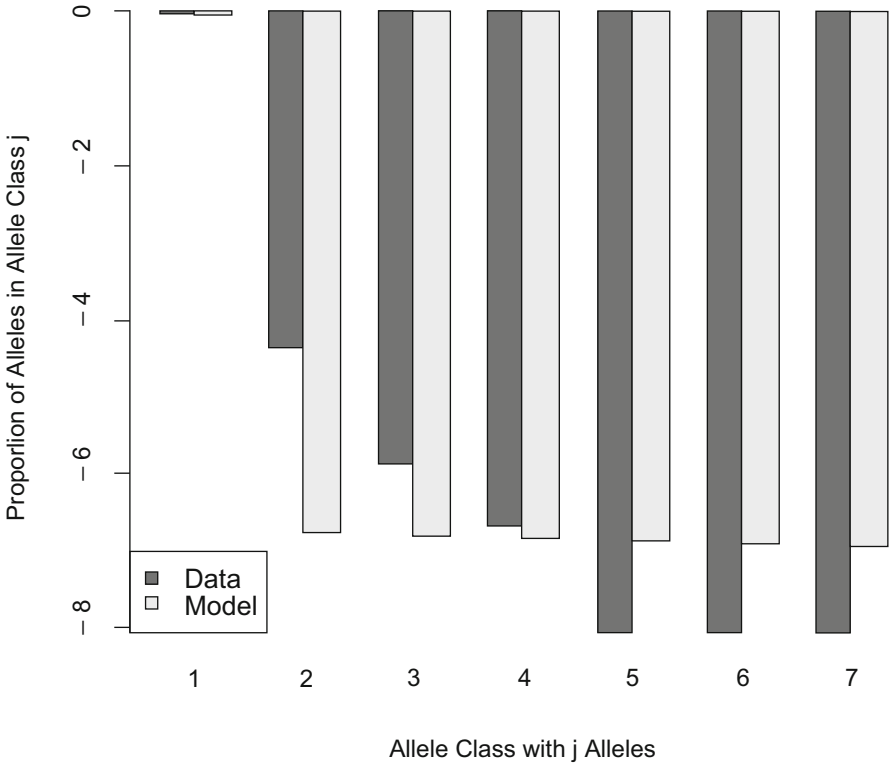


Fig. 7.15 AluYc1 data-based class frequencies against the theoretical $\{\Psi_k\}$ in log scale. Fitted by Griffiths–Pakes process with LF distribution, with $b = 0.035$, $p = 0.985$. (Source: Kimmel and Mathaes 2010b)

in Table 1 of Kimmel and Mathaes (2010b) suggest a high value for Ψ_1, we verify that the theoretical Ψ_1 attains such values for any choices of parameters b, p, and μ. For fixed $\mu = 10^{-6}$, we established a grid of b and p from 0 to 1 in steps of 0.01. Figure 7.11 shows that Ψ_1 can assume any value between 0 and 1, and that high values of Ψ_1 occur for a combination of low values of b and high values of p.

Comparison of Predictions to Observed Data

Kimmel and Mathaes (2010b) tested the theoretical distribution, based on the branching processes model, with the observed data of *Alu* sequences reported in the University of Southern California Human Genome database. For each of four different large (more than 1000 sequences) *Alu* subfamilies a consensus sequence was determined. The counts of *Alu* sequences belonging to a given subfamily that had n identical copies in the sample ($n = 1, 2, 3, \ldots, 7$) have been determined. These seven classes represent 99 % of the observed data. The observed distribution was

compared with the simulated distribution, using the χ^2 statistic. For simulations, the maximum liklihood method was used to estimate parameters b and p for each subfamily. The mutation rate was set to $\mu = 10^{-6}$, although the sensitivity to this parameter value was slight, with a range of 10^{-5}–10^{-9} per generation. Figures 7.12, 7.13, 7.14, and 7.15 depict the maximum-likelihood fits of the simulation data to the observed data for four *Alu* subfamilies. For each of the four subfamilies, the fit was good for classes 1 and 3, 4, 5, 6, and 7. But for each of the four subfamilies the fit for class 2 was poor. This difference in fit between simulations and observed data for class 2 is notable. It may be due to the method used to align sequences and recognize sequences in subfamilies, or to an unrecognized interesting biological process that has not been taken into account.

Chapter 8
Genealogies of Branching Processes and Their Applications

8.1 Genealogies of Branching Processes

One of the important questions in population dynamics and particularly in population genetics is how to gain information about a population's past, given its present status. Sources, historical in nature such as written records, archeological such as cemeteries, paleontological such as fossils, or even biological such as ancient DNA, are often of assistance. However, in many cases, all that is available is a sample from a contemporary population, with information about its demography or genetic makeup. Sometimes, a mathematical model of population growth may be assumed or statistically inferred from paleoecology or by other means. Human populations are of major interest, as are populations of endangered species. Other categories of biological genealogies are gaining prominence. Among them are genealogies of cells in cancerous tumors.

Two important approaches to population dynamics and genetics, which may be used for genealogical inference, are the Wright–Fisher model and the branching process model. Differences between these two will be described in the section concerning mitochondrial Eve (mtEve; Sect. 8.3). In this section, we will idiosyncratically review several approaches to modeling and inference based on branching process trees, as the Wright–Fisher model is covered exhaustively in the literature concerning coalescence (Wakeley 2009).

A comprehensive paper by Lambert (2008) provides an introduction to classical stochastic models of genealogies in discrete time, thus distinguishing models where the population size is fixed (models of Cannings, Wright–Fisher, and Moran) from the models where the population size fluctuates randomly (processes of Galton–Watson (GW) and birth–death processes). Continuous-time versions follow. One of the types of tools used in the theory are the jumping chronological contour processes (JCCP), which intuitively speaking is counting branch lengths corresponding to an exploration of the chronological tree (time pointing up in the Fig. 8.1a), then tracing back to the birth of a progeny branch (diagonally down), and so forth. Probabilistic properties of trees are equivalent to those of the contour processes, which frequently are random-walk-type processes, approximated by Brownian motions or state-continuous branching processes, given suitable scaling.

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0_8

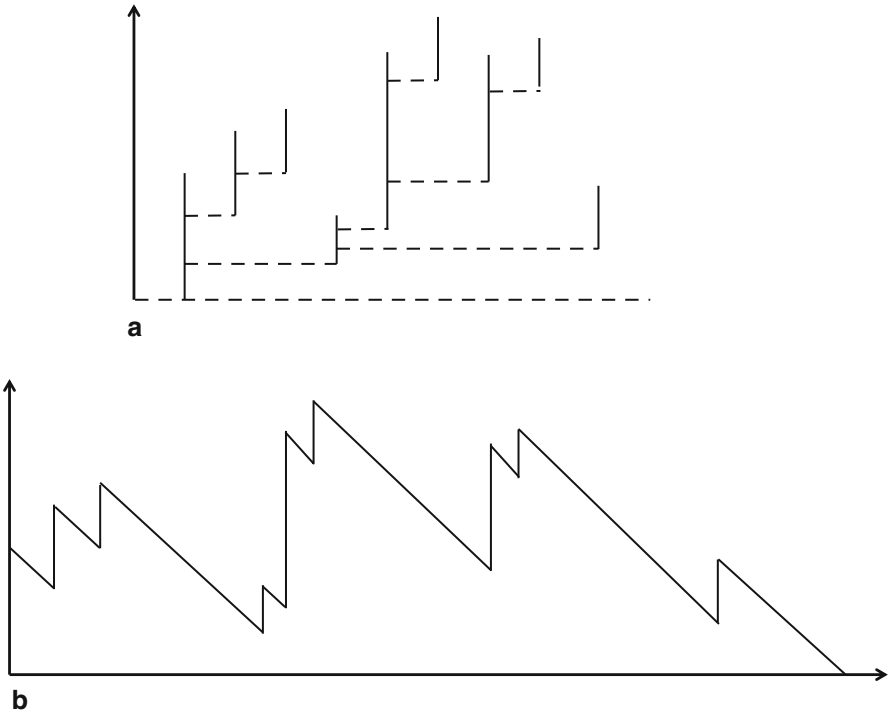


Fig. 8.1 A chronological tree a and the associated jumping chronological contour process (JCCP), with jumps in full line b. Please note that in part a vertical axes represents time. In part b, horizontal axes is a composite index of time and branching pattern, whereas the peaks of the graph correspond to times at death of the individual. (Modified after Lambert 2008)

Scaling limits of these models can be seen as genealogies of continuous populations. The continuous analogue of models with a fixed size is the stochastic flow on bridges of Bertoin and Le Gall (Le Gall 1999). The continuous analogue of branching models is the continuous-state branching process. Both processes have diffusion versions called, respectively, the Fisher–Wright diffusion and the Feller diffusion. Connections between the two kinds of models are also studied, and special attention is given to extinction/fixation (probability, expected time, conditioning). This is only a partial list of fundamental topics covered in Lambert (2008).

Another paper is Lambert and Popovic (2013), which defines a doubly infinite, monotone labeling of the GW genealogies. The genealogy of the current generation backwards in time is uniquely determined by the coalescent point process $(A_i, i \geq 1)$, where A_i is the coalescence time between individuals i and $i+1$. There is a Markov process of point measures $(B_i, i \geq 1)$ tracking ancestral relationships, such that A_i is also the first point mass of B_i. Following the analysis of the coalescent process, the paper is concerned with two applications in the discrete case. It is shown that the A_i's are iid when the offspring distribution is linear fractional. Another application concerns the law of Yaglom's quasi-stationary population size for subcritical Bienayme–Galton–Watson processes (BGW) processes.

An interesting review of mathematical properties of coalescence in branching trees can be found in a series of papers by Athreya and co-workers (Athreya 2012a, b). Another paper by one of the classics of the tree literature by David Aldous and co-workers, is Aldous et al. (2011), which concerns statistical questions related to diverse models of the "Tree of Life."

8.2 "Near-Critical" Processes

One of the interesting issues concerning a branching process is that of its genealogy. In broad terms, given a sample of individuals alive at given time t, we trace their past epochs of branching, i.e., the nearest common ancestors of the sample. This exercise is nontrivial since the sample we deal with consists of individuals with positively biased lifelengths. This latter effect is due to length-biased sampling analogous to that known from renewal processes. Our treatment is based on the paper by O'Connell (1995; also, see O'Connell 1993). It concerns the processes close to the critical process, in which the random effects are most pronounced. The theoretical results developed here are illustrated in Sect. 8.3 by an application concerning estimation of the age of the common female ancestor of modern humans.

We consider a family of time-continuous Markov branching processes (age-dependent processes with exponential lifetimes) parametrized by $t \geq 0$. Let $Z_t(u)$ be such a process with mean lifetime 1 and offspring distribution ξ_t with $\mathrm{E}\xi_t = 1 + \alpha/t + o(1/t)$ and $\mathrm{Var}(\xi_t) = \sigma^2 + o(1/t) < \infty$, where $\alpha \in R\backslash\{0\}$. We assume $\alpha \neq 0$ for notational convenience only; the corresponding results for the critical case can be extrapolated by letting $\alpha \to 0$. For this reason, we refer to it as the near-critical case. We will consider the genealogy of this process for fixed α and large t. A general reference on near-critical branching processes is the book of Jagers (1975), pp. 63–70, 199–206. We denote by $\mathrm{P}^x = \mathrm{P}_t^x$ the law of the process Z_t started at x, suppressing the subscript for notational convenience, and write E^x for the corresponding expectations. Set $f_t(s) = \mathrm{E}^1 s^{Z_t(1)}$. It is important to note (see, for example, Harris 1963) that the embedded (discrete-time) process $\{Z_t(n), n \in Z_+\}$ is a GW process with offspring mean $1 + \alpha/t + o(1/t)$, variance $\sigma^2 + o(1/t)$, and pgf $f_t(s)$. For $r > 0$, set $p_{x,r,t} = \mathrm{P}^x[Z_t(rt) > 0]$. We will assume throughout this section that $(Z_t^2(1)|Z_t(0) = 1)$ is uniformly integrable in t.

The first result describes the rate at which $p_{x,r,t} \to 0$ when $t \to \infty$, and an exponential limit law for near-critical Markov branching processes.

Theorem 8.1

1 *As* $t \to \infty$, $p_{x,r,t} \sim a_r x/t$, *where*

$$a_r = \frac{2\alpha}{\sigma^2}(1 - \mathrm{e}^{-\alpha r})^{-1}.$$

2 *If* $Z_t(0)/t \Rightarrow 0$ *as* $t \to \infty$, *then for* $\lambda > 0$, $x \in Z_+\backslash\{0\}$,

$$E^x[\exp\{-\lambda Z_t(rt)/t\}|Z_t(rt) > 0] \to \frac{b_r}{b_r + \lambda},\ t \to \infty,$$

i.e.,

$$P[Z_t(rt)/t\} > z|Z_t(rt) > 0] \to \exp(-b_r z),\ t \to \infty,$$

where $b_r = \mathrm{e}^{-\alpha r} a_r$. *The limit law is exponential with parameter* b_r.

The proof of the theorem is based on a direct diffusion approximation of the branching process.

The next result concerns the process reduced to individuals having living descendants. For each t and $0 \le u < t$, define the reduced process $N_t(u)$ to be the number of individuals in the process Z_t, alive at time u and having descendants alive at time t. Note that for each t, N_t is a time-inhomogeneous Markov branching process. In the statement of the theorem, $D_{Z_+}[0,1)$ denotes the space of càdlàg (continuous from the right, bounded from the left) paths in Z_+, parametrized to be defined on the unit interval; the weak convergence in this case requires only convergence of finite-dimensional distributions. The linear pure birth process with jump rate $b(t)$ is a time-continuous Markov chain $\{N(t), t \ge 0\}$, in which $\mathrm{P}[N(t + \Delta t) = N(t) + 1] = b(t)N(t) + o(\Delta t)$, where $o(\Delta t)/\Delta t \to 0$, when $\Delta t \to 0$.

Theorem 8.2 *As* $t \to \infty$, *the sequence of processes* $\{N_t(rt),\ 0 \le r < 1\}$ *converges in distribution in* $D_{Z_+}[0,1)$ *to a linear pure birth process* $\{N(r),\ 0 \le r < 1\}$ *with jump rate* $b(\alpha, r)N(r)$ *at time* r, *where*

$$b(\alpha, r) = \alpha(1 - \mathrm{e}^{-\alpha})^{-1}(1 - r)^{-1},$$

provided $N_t(0) \Rightarrow N(0)$. *In particular, as* $t \to \infty$,

$$P^x[N_t(rt) = k \mid N_t(0) = 1] \to q_r(1 - q_r)^{k-1},$$

where $q_r = [\exp(-r) - \exp(-\alpha)]/[1 - \exp(-\alpha)]$.

The result which is of most interest describes the degree of relationship of two randomly chosen individuals at time t. Let D_t denote the latest time, counting from the beginning of the process, at which the common ancestor of the two individuals exists. The following theorem provides the asymptotic distribution of this time:

Theorem 8.3 *For* $0 \le r < 1$, $x \in Z_+\backslash\{0\}$,

$$\lim_{t\to\infty} P[D_t > rt \mid N_t(0) = x] = \frac{2q_r^x}{(x-1)!}\{(x-1)!(q_r - 1)^{-x} - F(x-1, 1-q_r)\} \tag{8.1}$$

where $F : Z_+ \times (0,1) \to R$ *is defined by*

$$F(n, y) = \frac{\partial^n}{\partial y^n}\left\{\frac{\log(1-y)}{y^2}\right\}.$$

Proof The original proof in O'Connell (1995) is a modification of the corresponding result in Durrett (1978) for the critical case. The current proof is a slight modification of O'Connell (1995), which rectifies some inaccuracies in the original version of Expression (8.1). Let $P_{t,u,k}$ denote the probability that two individuals chosen randomly at time t have the same ancestor at time u, given $N_t(u) = k$. Let $X_1(u,t), \ldots, X_k(u,t)$ be independent and identically distributed random variables with the same distribution as $(Z_t(u)|Z_t(u) > 0)$. If we let

$$S_k(u,t) = X_1(u,t) + \cdots + X_k(u,t),$$

then

$$P_{t,u,k} = k\mathrm{E}\{[X_1(u,t)/S_k(u,t)]^2\}.$$

By Theorem 8.1, part 2, for each i and $0 \le r < 1$, $X_i(rt,t)$ converges in distribution, as $t \to \infty$, to an exponentially distributed random variable with mean b_r^{-1}, which we denote by $X_i(r)$. So by bounded convergence we have

$$P_{t,rt,k} \to k\mathrm{E}\{[X_1(r)/S_k(r)]^2\},$$

as $t \to \infty$, where

$$S_k(r) = X_1(r) + \cdots + X_k(r).$$

Random variable $X_1(r)/S_k(r)$ can be represented as $Z = X/(X+Y)$, where $X \sim \exp(\psi)$ and $Y \sim$gamma$(\psi, k-1)$, and X and Y are independent. Independent of constant ψ, this ratio has distribution with density $f_Z(z) = (k-1)(1-z)^{k-2}$, $z \in [0,1]$, and consequently, $\mathrm{E}(Z)^2 = 2/[k(k+1)]$. This yields

$$\mathrm{E}[kX_1(r)/S_k(r)]^2 = \frac{2k}{k+1}.$$

Combining this with Theorem 8.2 we have as $t \to \infty$,

$$\begin{aligned}
\mathrm{P}[D_t > rt \mid N_t(0) = x] &= \sum_{k=1}^{\infty} P_{t,rt,k}\mathrm{P}[N_t(rt) > k \mid N_t(0) = x] \\
&\to \sum_{k=1}^{\infty} \frac{2}{k+1}\mathrm{P}[N(r) > k \mid N(0) = x] \\
&\to \sum_{k=1}^{\infty} \frac{2}{k+1}\binom{k-1}{x-1} q_r^x (1-q_r)^{k-x},
\end{aligned}$$

However, by the definition of $F(y,n)$, we have

$$\begin{aligned}
F(y,n) &= \frac{\partial^n}{\partial y^n}\left[\frac{\ln(1-y)}{y^2}\right] = -\sum_{k\ge 1}\frac{\partial^n}{\partial y^n}\left(\frac{y^{k-2}}{k}\right) \\
&= \frac{(-1)^{n+1}n!}{y^{n+1}} - \sum_{k\ge 2}\frac{(k-2)!}{(k-2-n)!(k+1)}y^{k-2-n},
\end{aligned}$$

and consequently,

$$F(1-q_r, x-1) = (x-1)!(q_r-1)^{-x} - \sum_{k\geq 1} \frac{(k-1)!}{(k-x)!(k+1)}(1-q_r)^{k-x},$$

and the result follows.

Remarks

1. The limiting process in Theorem 8.2 can be represented as a deterministic time change of a (time-homogeneous) Yule process (in this case, a Markov age-dependent branching process, with progeny number equal to two and life-length being a random variable distributed exponentially with parameter λ). If $\{Y(t), t \geq 0\}$ is a Yule process with branching rate 1, then the process $\{Y[\ln((1-e^{-\alpha})(e^{-r\alpha}-e^{-\alpha}))],\ 0 \leq r < 1\}$ has the same law as N.
2. It is instructive to derive explicit expressions for O'Connell's (1995) limit distributions $\Phi_x(r) = \lim_{t\to\infty} P[D_t/t > r \mid N_t(0) = x]$. We obtain

$$\Phi_1(r) = \frac{2q_r}{(1-q_r)^2}(q_r - 1 - \ln q_r),$$

$$\Phi_2(r) = \frac{2q_r}{(1-q_r)^3}(1 - q_r^2 - 2q_r \ln q_r),$$

   where $r \in [0,1]$. Let us note that $\Phi_1(1) = \Phi_2(1) = 0$, but $\Phi_2(0) = 2/3$ while $\Phi_1(0) = 1$. The reason is that in case of the process started by $x = 2$ ancestors, there is a probability equal to $1/3$ that two randomly selected descendants are traced to different ancestors.
3. Similar and related results for general branching processes can be found in Sagitov (1989), Taïb (1987), and Zubkov (1975), for branching diffusion processes in Durrett (1978) and Sawyer (1976), and for superprocesses in Dawson and Perkins (1991) and Etheridge (1992). For an excellent review of the vast literature on genealogical processes in population genetics models, see Tavaré (1984).

8.3 Application: Estimation of the Age of the Mitochondrial Eve

8.3.1 Population Genetic Model

One of the applications of O'Connell's (1995) results is the estimation of the age of the process, which is not observable, based on statistics describing the ages of most recent common ancestors (MRCA) of the pairs of extant (contemporary) individuals sampled from the process. The time from MRCA of two individuals can be measured using divergence (mismatch) between DNA sequences ascertained in these individuals. O'Connell (1995) presents such an analysis leading to an estimate of the time when the female ancestor of modern humans (mtEve) lived. We provide

an account of his methodology. The estimates obtained differ from those obtained using more accepted methods like the coalescence theory (Griffiths and Tavaré 1999). However, the originality of O'Connell's (1995) approach is sufficient to justify this presentation.

Wilson and Cann (1992) and Vigilant et al. (1989, 1991) were the first to hypothesize that a female ancestor of modern humans probably lived in Africa about 200,000 years ago. Hasegawa and Horai (1990) estimate the age to be equal to 280,000 years. Stoneking et al. (1992) published an estimate of 135,000 years. For other more recent examples of estimation of past demographic trends, see e.g., Harpending et al.(1998) and Kimmel et al. (1998).

The data used by O'Connell (1995) is a collection of aligned nucleotide sequences, each approximately 600 base pairs (sites) in length, sampled from the hypervariable segment in the control region of the human mitochondrial genome, of 189 individuals from around the world. There are four nucleotides: adenine, thymine, guanine, and cytosine. A typical sequence might be coded as follows:

$$TTCTTTCCATGGGGAAGCAGA \cdots CCTAACCAGA.$$

It is accepted that mitochondrial sequences are maternally inherited and that mitochondrial DNA (mtDNA) mutations are neutral, or very rare, from the standpoint of natural selection. In other words, the specific makeup of mtDNA does not influence individual's reproductive fitness.

The following model of mutation is known as the infinite sites model (ISM). A substitution occurs if one of the nucleotides in the sequence is replaced by another, and the new sequence is inherited. According to the molecular clock hypothesis substitutions occur randomly along lineages at a constant rate, and rates along different lineages are the same. The genetic distance, or divergence, between two such sequences is defined to be the proportion of sites at which the sequences differ. Among humans this is typically less than 5 % in the control region of mtDNA. Vigilant et al. (1989) found the average divergence between the humans in their sample, and a sample chimpanzee, to be about 15 %. The divergence rate is very small, so over the time period we are considering here (the post-Eve period), we can assume that each substitution produces a new type, that is, reverse substitutions do not occur. Thus, if the MRCA of two individuals died u million years ago, the number of differences between their mtDNA types will be a random variable with distribution approximated by the Poisson distribution with mean $2u\mu$, where μ is the substitution rate (in units of number of substitutions per million years).

Now suppose two individuals are sampled randomly from the current population, and δ denotes the rate of divergence (in units of percentage divergence per million years). Note that if l denotes the sequence length, then $\delta = 2\mu/l$. If we have a model for the genealogical structure of the population, then the expected amount of divergence between the mtDNA sequences of the two individuals will be equal to the expected time back to the common ancestor of the two individuals (under our model, in units of millions of years), multiplied by the divergence rate, δ.

Assume that the female population size follows a Markov branching process Z_T with mean number of offspring $1 + \alpha/T$, where $T = T_a/\lambda$; T_a is the time to the

MRCA (T_{MRCA}), λ is the mean effective lifetime (or generation time) and $\alpha \in R$ is the growth parameter.

To obtain an indication of how fast the population might have been growing, suppose the estimate of 200,000 years were correct. Then a straightforward moment calculation based on this model with offspring variance $\sigma^2 = 2$, mean (effective) lifetime 25 years, and current (effective) female population size 1 billion, yields the rough estimate $\hat{\alpha} = 13.7$ (c.f. Eq. 8.2 below). The estimate $\hat{\alpha}$ is quite insensitive to apparently large adjustments in these values, and remains in the "slightly supercritical" framework for quite a wide range.

We will slightly depart from the original method of estimation in O'Connell (1995). We will use the process with the single ancestor, Eve, i.e., with $x = 1$, while O'Connell (1995) used processes generated by the (almost surely) two direct descendants of Eve ($x = 2$). The results are almost identical and our method seems simpler. If we start time at the birth of Eve, then $N_T(0) = 1$. Then $Z_T(T)$ is the current (effective) female population size. Using the approximation results in Theorems 8.1–8.3, we can simultaneously estimate α and T, based on the observations $Z_T(T)$ and the average pairwise divergence in a random sample of n contemporary individuals $\bar{d}_n$. We will assume for the moment that the divergence rate δ is known. Denote by λ the mean effective lifetime of an individual. By Theorem 8.1, part 2,

$$\mathrm{E}[Z_T(T) \mid N_T(0) = 1] \simeq \frac{T_\alpha/\lambda}{b_r} = \frac{\sigma^2 T_\mathrm{a}}{2\lambda\alpha}\left(\mathrm{e}^\alpha - 1\right). \tag{8.2}$$

We also have, by Theorem 8.3,

$$\begin{aligned} \mathrm{E}[\bar{d}_n \mid N_T(0) = 1] &\simeq \delta\lambda \mathrm{E}[T - D_T \mid N_T(0) = 1] \\ &= \delta T_\mathrm{a}\left\{1 - \textstyle\int_0^1 \mathrm{P}[D_T > rT \mid N_T(0) = 1]dr\right\} \\ &\simeq \delta T_\mathrm{a}\gamma_1(\alpha), \end{aligned} \tag{8.3}$$

where

$$\gamma_1(\alpha) = 1 - \int_0^1 \Phi_1(r)dr = 1 - 2\int_0^1 \frac{q_r}{(1-q_r)^2}(q_r - 1 - \ln q_r)dr. \tag{8.4}$$

One can simplify (8.4) to get

$$\gamma_1(\alpha) = 1 - 2\alpha^{-1}\int_0^1 \frac{v}{(1-v)^2(v+\kappa)}(v - 1 - \ln v)\mathrm{d}v, \tag{8.5}$$

where

$$\kappa = \frac{\mathrm{e}^{-\alpha}}{1 - \mathrm{e}^{-\alpha}}. \tag{8.6}$$

Note that $\gamma_1(\alpha)$ is positive and increase in α, and $\gamma_1(\alpha) \uparrow 1$ as $\alpha \to \infty$. For the simplest moment-based estimates, assuming that δ, σ^2, and λ are known, just set

Table 8.1 Estimates of the parameters of O'Connell (1995) model

$\lambda Z_T(T)/\sigma^2$	δ	$\hat{\alpha}$	$\hat{T}_a/10^3$
12.5×10^9	1.8	12.062	1741
	2.7	12.508	1156
	4	12.939	777
5×10^9	1.8	11.047	1761
	2.7	11.497	1168
	4	11.932	785
30×10^9	1.8	13.023	1726
	2.7	13.465	1147
	4	13.893	772

$$Z_T(T) = \frac{\sigma^2 \hat{T}_a}{\lambda \hat{\alpha}} (e^{\hat{\alpha}} - 1), \tag{8.7}$$

$$\hat{T}_a = \frac{\bar{d}_n}{\delta \gamma_1(\hat{\alpha})}, \tag{8.8}$$

and solve for $(\hat{\alpha}, \hat{T}_a)$. Although σ^2 is unknown, when α is sufficiently large, the actual value (within reason) will not affect the estimates considerably. (This is due to the dominating exponential term in Eq. 8.7.) The same is true for λ.

Remarks concerning performance of the estimators can be found in O'Connell (1995).

8.3.2 Numerical Estimates

Of the 189 individuals considered by Vigilant et al. (1989), O'Connell (1995) picked a subsample of 19, without being deliberately biased in any way. The sample consists of six Asians, one native Australian, one Papua New Guinean, six Europeans and five Africans. A histogram of the 171 pairwise differences in this sample is shown in Fig. 4 of O'Connell (1995). The average divergence was found to be 2.8 %.

In June 1992, according to the Population Reference Bureau estimates, the human population size was approximately 5.412 billion. This gives about 1 billion as a rough estimate for the 1992 effective female population size, assuming that about half the population is female, and that the current female population represents approximately 2.7 generations. The estimates are quite insensitive to variations in this figure.

Note that the estimates $\hat{\alpha}$ and $\hat{T}_a$ are functions of $\lambda Z_T(T)/\sigma^2$ and δ; these are shown in Table 8.1, for various different values of $\lambda Z_T(T)/\sigma^2$ and δ.

These estimates differ only slightly from the original O'Connell's (1995) numbers. If $Z_T(T) = 1$ billion, $\sigma^2 = 2$ and $\lambda = 25$, then $\lambda Z_T(T)/\sigma^2 = 12.5$ billion. Although

these choices seem somewhat arbitrary, we can see from the Table 8.1 that any kind of realistic deviations from these values will have little or no effect on the estimates. The most important parameter is δ, the rate of divergence.

To derive the estimates for the growth rate, $\hat{\alpha}$, and the age of Eve, $\hat{T}_{\mathrm{a}}$, we simply calculated the expected current population size and the expected average pairwise divergence in a sample of contemporary individuals, and assumed the other parameters were known. We are therefore not fully utilizing the information contained in the sample. It might be helpful to know more about the joint distribution of the pairwise divergences (d_{ij}), or the joint distribution of the respective frequencies of distinct types, in a finite sample. The latter would be analogous to Ewens' sampling formula for the infinite alleles Wright–Fisher model for neutral evolution (Nagylaki 1990), which is not applicable to the Eve problem because it is based on the assumption that the population size is constant over time.

8.3.3 Robustness of Mitochondrial Eve

The model originally proposed by O'Connell (1993, 1995) has limit results based on the assumption that the population is growing as a slightly supercritical GW branching process. It is interesting to compare the results of that branching process model with results of two other well-known models of population genetics, the Wright–Fisher model and the coalescence-based model (Barton 2007, Chap. 15).

The Wright–Fisher model for genetic drift assumes that:

- The population size is constant.
- The generations are nonoverlapping.
- There is random mating between individuals with different alleles.
- Alleles are neutral to selection.
- There is random sampling.
- No additional mutant alleles occur as the population evolves.

The Wright–Fisher model views changes among individuals as a population evolves forward from an initial population.

In contrast, the coalescent-based model views a population backward. A sample of differing individuals in the final population is analyzed, and this information, assuming no selection, is used to infer their MRCA and the time (in generations) to the MRCA (T_{MRCA}).

These models can be applied to the problem of human population trajectory using mtDNA sequence data. These data include the hypervariable region of a Neanderthal mitochondrial genome (Krings et al. 1999), the complete mitochondrial genome of one Neanderthal (Green et al. 2008), the complete mitochondrial genome of five Neanderthals (Briggs et al. 2009), and a sample of 663 mitochondrial sequences of modern humans (Krings et al. 1999). These data provide an opportunity to study the sensitivity of genetic variation indices to departures from the assumptions made in different models. In particular, it is interesting how these departures influence

the distributions of the time to coalescence. The expected values for the T_{MRCA}, and other parameters, can be computed for each of the three models: the Wright–Fisher model, the coalescent-based model, and the O'Connell model. Results are compared to full genealogies generated by computer simulations that are based upon a slightly supercritical GW branching process (Cyran and Kimmel 2010).

The Wright–Fisher model has been applied to the smallest sample exhibiting effects of the genetic drift, i.e., the sample composed of two DNA sequences. The model assumes a population of haploid individuals, say mtDNA sequences, which at time $t = 0$ has the size N_t. Since multinomial sampling is assumed, two individuals at generation $t+1$ are descendants of the single member of generation t with probability $p_t = 1/N_t$. Consequently, with probability $q_t = 1 - p_t$ they are descendants of two different members. This is reflected in the following distribution of the time to coalescence, T_{2c}, of two randomly drawn chromosomes (Bobrowski and Kimmel 2004):

$$P(T_{2c} = t) = \prod_{k=T-t}^{T-1} q_k - \prod_{k=T-t-1}^{T-1} q_k = p_{T-t-1} \prod_{k=T-t}^{T-1} q_k, \tag{8.9}$$

where T denotes the number of generations we consider, and for mathematical consistency, we let $q_{-1} = 0$ and $p_{-1} = 1$. The average pairwise mutation difference within a sample, after scaling by the mutation rate μ, corresponds to the expectation of the coalescence distribution (Bobrowski and Kimmel 2004), and moreover, the discrete nature of generations makes it easy to simulate the demography. Therefore, using Monte Carlo techniques, it is possible to estimate the unconditional coalescence distribution by averaging the conditional one, using a series of N_t realizations required to compute parameters in Eq. (8.9).

For the coalescent model, it is assumed that population size N_τ is variable in time and continuous time τ is measured backwards. Suppose also that $\lambda(\tau) = N_0/N_\tau$ and that τ_{2c} is the time to coalescence of a pair of chromosomes observed over N_0 generations. Then, the tail of the distribution of τ_{2c} is given by

$$P(\tau_{2c} > \tau) = \exp\left[-\int_0^\tau \lambda(u)\mathrm{d}u\right], \tag{8.10}$$

which is the continuous analog of Eq. (8.9). To ensure existence of a unique common ancestor, $\lambda(t)$ must satisfy

$$\int_0^\infty \lambda(u)\mathrm{d}u = \infty. \tag{8.11}$$

For the stochastic N_t, and therefore $\lambda(\tau)$, the right side of the Eq. (8.10) should be averaged over the process realizations.

It is worth noticing that the continuous coalescence model correctly approximates the discrete coalescent model as long as $1 - 1/N_\tau \approx \exp(-1/N_\tau)$, which certainly is not true in the early phase of the branching process, when N_t is not large and

undergoes stochastic fluctuations. This fact is reflected in the differences between experimental distributions of the time to coalescence in the coalescent model and the O'Connell branching process model.

The O'Connell model assumes a slightly supercritical time-homogenous branching process with expected number of offspring $E(\zeta_0) = 1 + \alpha/T + o(1/T)$ and variance $\mathrm{Var}(\zeta_0) = \sigma^2 + o(1/T)$. For this model, the asymptotic of the probability $P^x(Z_t > 0)$, where P^x denotes probabilities for the process started by x individuals, satisfies the O'Connell (1995) formula

$$P^x(Z_t > 0) \sim \frac{2\alpha x}{\sigma^2\left[1 - \exp\left(-\alpha\frac{t}{T}\right)\right]}, T \to \infty. \tag{8.12}$$

From this it follows (Cyran and Kimmel 2004) that

$$E(Z_T | Z_T > 0, Z_0 = x) \sim \frac{\sigma^2\, T}{2\alpha}\left[\exp(\alpha) - 1\right], T \to \infty, \tag{8.13}$$

where the symbol $\sim$ denotes asymptotic equivalence. The time interval $[0, T]$ of variable t is expressed as a unit interval [0,1] of variable $r = t/T$. Then based on Theorems 8.2 and 8.3 for times T long enough, the following equation describes the tail of the distribution of D_T, the time of death of the last common ancestor of two individuals living at T, given that the population history is started from x individuals having descendents at T:

$$\lim_{T\to\infty} P\left(\frac{D_T}{T} > r | K_0 = x\right) = \frac{2q_r^x}{(x-1)!}\left[(q_r - 1)^{-x}(x-1)! - F(x-1, 1-q_r)\right], \tag{8.14}$$

where

$$q_r = \frac{\mathrm{e}^{-r\alpha} - \mathrm{e}^{-\alpha}}{1 - \mathrm{e}^{-\alpha}} \tag{8.15}$$

and $F : \mathbb{Z}_+ \times (0,1) \to \mathbb{R}$ is defined as

$$F(n, y) = \frac{\partial^n}{\partial y^n}\left[\frac{\ln(1-y)}{y^2}\right]. \tag{8.16}$$

The O'Connell original distribution is continuous, but to compare it to the discrete empirical distributions the discretized version is specified by the tail of the original distribution computed at points r corresponding to integer values of $t = rT$. For the sake of terminological simplicity, this discrete distribution is still referred to as the O'Connell distribution.

In the O'Connell model,

$$E\left(\frac{T_{2c}}{T} | K_0 = 1\right) = \frac{1}{T} E\left[(T - D_T) | K_0 = 1\right], \tag{8.17}$$

and

$$\hat{T}_{MRCA(y)} = E\left(\frac{T_{MRCA}}{T}|K_0 = 1\right) \times \frac{d_{avg}}{\delta\left(1 - 2\int_0^1 \frac{\hat{q}_r}{(1-\hat{q}_r)^2}(\hat{q}_r - 1 - \ln\hat{q}_r)\mathrm{d}r\right)}, \tag{8.18}$$

where

$$\hat{q}_r = \frac{\mathrm{e}^{-r\hat{\alpha}} - \mathrm{e}^{-\hat{\alpha}}}{1 - \mathrm{e}^{-\hat{\alpha}}}, \tag{8.19}$$

and $T_{\mathrm{MRCA}(y)} = \lambda T_{MRCA}$ is the T_{MRCA} expressed in years.

Moreover, the expectation of the ratio T_{MRCA}/T in Eq. (8.18) can be obtained from simulations with recorded full genealogies (Cyran and Kimmel 2010). Therefore, to calculate the estimated MRCA time $\hat{T}_{MRCA}$ from genetic variation data, $\hat{\alpha}$ is needed. However, from simulation results concordant with limiting properties of the O'Connell model $E(T_{MRCA}/T|K_0 = 1)$. Therefore, $T_{MRCA(y)}$ and α can be simultaneously estimated. From Eq. (8.13), if Z_T is substituted as an estimate of its expected value, it follows that

$$Z_T = E\left(\frac{T}{T_{MRCA}}|K_0 = 1\right)\frac{\sigma^2\hat{T}_{MRCA(y)}}{2\lambda\hat{\alpha}}\left[\exp(\hat{\alpha}) - 1\right] \tag{8.20}$$

and estimates of $T_{MRCA(y)}$ and α are solutions of the system of Eqs. (8.18) and (8.20) for given effective population size of females Z_T, and genetic data d_{avg} and δ, where d_{avg} is the average pairwise mutation difference among sequences in the sample and $\delta = \mu/\lambda$ is the divergence rate, with μ being the mutation rate and λ being the mean effective lifetime of an individual.

The expected values of $T_{MRCA(y)}$ of modern humans calculated for the three models are the following: Wright–Fisher 168–189,000, coalescent 165–187,000, and O'Connell limit 176,000 years. See Cyran and Kimmel (2010) for details.

These results indicate that the estimates of the time to coalescence in Wright–Fisher and the coalescent models with random population size are quite robust to the model assumptions. They deviate by less than 8 % from the O'Connell model predictions, and the asymptotic O'Connell prediction differs from the actual value computed in the full genealogy model by only 1.6 % (Cyran and Kimmel 2010).

Appendix A
Multivariate Probability Generating Functions

In this section, we will collect some results, which are referred to throughout the book. Suppose $X = (X_1, \ldots, X_n) \sim \left\{p_{i_1 i_2 \ldots i_n}\right\}_{i_1, i_2, \ldots, i_n \geq 0}$ is a finite vector of non-negative random variables, or a Z_+^n-valued rv.

Definition A.1 Definition of the multivariate pgf. *The pgf f_X of a Z_+^n-valued rv X is function*

$$f_X(s) = \mathrm{E}\left(s_1^{X_1} s_2^{X_2} \ldots s_n^{X_n}\right) = \sum_{i_1, i_2, \ldots, i_n \geq 0} p_{i_1 i_2 \ldots i_n} s_1^{i_1} s_2^{i_2} \ldots s_n^{i_n} \tag{A.1}$$

well defined if $s = (s_1, s_2, \ldots, s_n) \in U_n \equiv [0, 1]^n$.

Theorem A.1 Multivariate pgf Theorem. *Suppose X is a Z_+^n -valued rv with pgf f_X. Let us denote (N_i) the nontriviality condition for the i-th coordinate of X, $P[X_i \leq 1] < 1$.*

1. *f_X is nonnegative and continuous with all derivatives. Under (N_i), it is increasing and convex as a function of s_i.*
2. *The marginal laws for subsets of X_i's can be obtained by setting respective arguments of the pgf equal to 1, e.g., $f_X(s)_{|s_j=1,\ j\neq i} = f_{X_i}(s_i)$, etc.; $f_X(e) = 1$, where $e = (1, \ldots, 1)$.*
3. *$\partial^{k_1+\ldots+k_n} f_X(0)/\partial s_1^{k_1} \ldots \partial s_n^{k_n} = k_1! \ldots k_n!\ p_{k_1 \ldots k_n}$.*
4. *The $(k_1, \ldots, k_n)$-th mixed factorial moment of X, $\mu_{k_1,\ldots,k_n} = E\left[\prod_{i=1}^{n} \prod_{j=0}^{k_i - 1} (X_i - j)\right]$, is finite if and only if $\partial^{k_1+\ldots+k_n} f_X(e-)/\partial s_1^{k_1} \ldots \partial s_n^{k_n} = \lim_{s \uparrow e} \partial^{k_1+\ldots+k_n} f_X(s)/\partial s_1^{k_1} \ldots \partial s_n^{k_n}$ is finite. In such case $\mu_{k_1,\ldots,k_n} = \partial^{k_1+\ldots+k_n} f_X(e-)/\partial s_1^{k_1} \ldots \partial s_n^{k_n}$.*
5. *If X and Y are two independent Z_+^n-valued rv's, then $f_{X+Y}(s) = f_X(s)\ f_Y(s)$.*

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology*,
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

6. *If Y is a Z_+^n-valued rv and $\left\{X_j^{(i)};\ i \geq 1\right\}$, $j = 1, \ldots, n$ are sequences of Z_+^m-valued rv's, then $V = \sum_{j=1}^{n} \sum_{i_j=1}^{Y_j} X_j^{(i_j)}$ is a Z_+^m-valued rv with pgf $f_V(s) = f_Y\left[f_{X_1^{(1)}}(s), \ldots, f_{X_n^{(1)}}(s)\right]$, $s \in U_m$.*
7. *Suppose $\{X_i;\ i \geq 1\}$ is a sequence of Z_+^n-valued rv's. The limit $\lim_{i\to\infty} f_{X_i}(s) = f_X(s)$ exists for each $s \in U^n$ if and only if the sequence $\{X_i;\ i \geq 1\}$ converges in distribution, i.e., when $\lim_{i\to\infty} P[X_{i,1} = k_1, \ldots, X_{i,n} = k_n] = P[X_1 = k_1, \ldots, X_n = k_n]$. Then $f_X(s)$ is the pgf of the limit rv X.*

A further generalization to the denumerable infinite case is possible. Suppose $X = (X_1, \ldots, X_n, \ldots) \sim \left\{\left\{p_{i_1 i_2 \ldots i_n}\right\}_{i_1, i_2, \ldots, i_n \geq 0}\right\}_{n \geq 1}$ is an infinite vector of non-negative random variables, with the σ-algebra generated by the finite-dimensional truncations of the sequence. Also, we may consider X a Z_+^∞-valued rv.

Definition A.2 Denumerable pgf definition. *The pgf f_X of a Z_+^∞ -valued rv X is a function*

$$f_X(s) = \mathrm{E}\left(s_1^{X_1} s_2^{X_2} \ldots s_n^{X_n} \ldots\right) = \sum_{i_1, i_2, \ldots, i_n \geq 0} p_{i_1 i_2 \ldots i_n} s_1^{i_1} s_2^{i_2} \ldots s_n^{i_n} \tag{A.2}$$

defined for

$$s \in \bigcup_{n \geq 1} U_n \equiv \bigcup_{n \geq 1} \{(s_1, s_2, \ldots, s_n, 1, 1, \ldots) : s_1, s_2, \ldots, s_n \in [0, 1]\}, \tag{A.3}$$

i.e., for arguments $s \in [0, 1]^\infty$ with only finite number of coordinates not equal to 1.

Properties 1 through 5 stated in the multivariate pgf Theorem carry over to the finite-dimensional restrictions of the denumerable pgf. Important difference is that Property 6 does not necessarily hold for infinite n, since the resulting sum may be improper (if it is proper, then Property 6 holds). Also the convergence Property 7 requires an additional continuity requirement:

Denumerable pgf Convergence Suppose $\{X_i, i \geq 1\}$ is a sequence of Z_+^∞-valued rv's. A necessary and sufficient condition for convergence in distribution of this sequence to a Z_+^∞-valued rv X is that $\lim_{i\to\infty} f_{X_i}(s) = f_X(s)$ exists for each $s \in \bigcup_{n \geq 1} U_n$, and that f_X is pointwise continuous for all sequences $\{s^{(i)}, i \geq 1\}$ with $s^{(i)} \in U_n$. Then $f_X(s)$ is the pgf of the limit rv X.

Appendix B
Probability Distributions for the Bellman–Harris Process

B.1 Construction

We start with a rigorous construction of the probability space of the process, following Chap. 6 of Harris (1963). The elements of the probability space are *family histories* of the particles.

B.1.1 The Families

Let $\mathcal{I}$ be the collection of elements ι, where ι is either 0 or a finite sequence of positive integers $i_1, i_2, \ldots, i_k$. The collection $\mathcal{I}$ is denumerably infinite. The elements ι are enumerated in a sequence $\iota_1, \iota_2, \ldots$, starting for example with $0, 1, 2, 11, 3, 21, 12, 111, \ldots$. The *ancestor* or *founder* is denoted by $< 0 >$, while $< i_1, i_2, \ldots, i_k >$ denotes the i_k-th child of the i_{k-1} -th child of $\ldots$, of the i_2-th child of the i_1-th child of the ancestor.

The *family history* ω is the sequence $\omega = (l, \nu; l_1, \nu_1; l_2, \nu_2; l_{11}, \nu_{11}; \ldots)$ where l_ι is a nonnegative real and represents the length of life of ι, while ν_ι is a nonnegative integer and represents the number of children of ι. The collection of all family histories is denoted by Ω. Family history is a *redundant* description of the particles pedigree in the sense that it enumerates even "nonexistent" children; for example if $\nu_{ij} = 5$ (the j-th child of the i-th child of the ancestor has five children), then none of the pairs l_{ijk}, ν_{ijk} for $k > 5$ corresponds to any members of the pedigree.

For each $\omega \in \Omega$, we define a sequence $I_0(\omega), I_1(\omega), \ldots$, where I_k is a collection of objects $< \iota >$ called the k-th *generation*. The 0-th generation $I_0(\omega)$ is the ancestor $< 0 >$, and $I_1(\omega)$ is the set of all objects $< i >$ with $1 \le i \le \nu(\omega)$. The succeeding generations are defined inductively: $I_k(\omega)$ is the set of all objects $< i_1 i_2 \ldots i_k >$ such that $< i_1 i_2 \ldots i_{k-1} >$ belongs to $I_{k-1}(\omega)$ and $i_k \le \nu_{i_1 i_2 \ldots i_{k-1}}(\omega)$. The set of objects $\bigcup_{k=0}^{\infty} I_k(\omega)$ is called the *family* $I(\omega)$. In view of remarks in the preceding paragraph, more than one *family history* ω may, in general, correspond to the same *family* $I(\omega)$.

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology*,
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

B.1.2 The Number of Objects at Given Time

If the object $< \iota >=< i_1 \dots i_k >$ belongs to the family $I(\omega)$, it is born at the time $t' = l + l_{i_1} + \dots + l_{i_1 i_2 \dots i_{k-1}}$ and dies at the time $t'' = t' + l_{i_1 i_2 \dots i_k}$ if $t \in [t', t'')$, then the age of the object at t is $t - t'$. Thus, if at time t we count the objects that are alive and have ages $\le y$, then $< \iota >$ is counted if and only if the following conditions hold (with obvious modifications if $\iota = 0$):

$$\left.\begin{aligned} i_1 \le \nu, i_2 \le \nu_{i_1}, \dots, i_k \le \nu_{i_1 i_2 \dots i_{k-1}}, \\ t - y \le l + l_{i_1} + \dots + l_{i_1 i_2 \dots i_{k-1}} \le t, \\ l + l_{i_1} + \dots + l_{i_1 i_2 \dots i_{k-1}} + l_{i_1 i_2 \dots i_k} > t. \end{aligned}\right\} \tag{B.1}$$

The first line in (B.1) means that $< \iota >$ belongs to the k-th generation; the second line says that $< \iota >$ was born between $t - y$ and t; the third line says that $< \iota >$ dies after time t.

For each object ι let us define $Z_\iota(y, t, \omega)$ to be 1 if (B.1) holds and to be 0 otherwise. Define

$$Z(y, t, \omega) = \sum_{\iota \in \mathcal{I}} Z_\iota(y, t, \omega)$$

and

$$Z(t, \omega) = Z(\infty, t, \omega) = \sum_{\iota \in \mathcal{I}} Z_\iota(\infty, t, \omega).$$

Thus $Z_\iota(y, t, \omega)$ is 1 if $< \iota >$ is alive and of age $\le y$ at t and 0 otherwise; $Z(y, t, \omega)$ is the total number of objects of age $\le y$ at t; and $Z(t, \omega)$ is the total number of objects at t. The possibility $Z(y, t, \omega) = \infty$ for some values of y, t, ω is admitted.

Let us note that if $Z(t_0, \omega_0) = 0$ for some t_0, ω_0, then $Z(t, \omega_0) = 0$ for all $t > t_0$.

B.1.3 Probability Measure

Definition B.1 *The probability measure* P *is built on the space* Ω *of family histories* ω *in the following way:*

1. *The random variables* l_ι *are iid with distribution*

$$\mathrm{P}\{l_\iota \le t\} = G(t),$$

where G *is a right-continuous probability distribution function for which* $G(0+) = 0$.

2. *The ν_t's are independent of each other and of the l's, and id with the pgf*

$$f(s) = \sum_{r=0}^{\infty} p_r s^r = \sum_{r=0}^{\infty} \mathrm{P}\{\nu_t = r\} s^r,$$

with the trivial cases excluded and $m \equiv f'(1-) < \infty$.

We denote the k-th convolution of G with itself by G^{*k} ($G^{*1} = G$). Thus

$$G^{*k}(t) = \int_{0-}^{t+} G^{*(k-1)}(t-y) d_y G(y).$$

Since ω corresponds to a denumerable family of independent real-valued random variables, the basic theorem of Kolmogorov insures that the above assumptions determine uniquely a countably additive probability measure P on the σ -algebra generated by the cylinder sets in Ω. From the definition of $Z(t,\omega)$, it is seen that Z is measurable in (t,ω), where the measurable (t,ω) sets are those generated by rectangles $A \times B$, A being a Borel t-set and B a measurable set in Ω. This conclusion is equivalent to a statement that the family of rv's $\{Z(t,\omega), t \geq 0\}$ is a *stochastic process*.

B.1.4 The Embedded Galton–Watson Process and Extinction Probability

Let $\zeta_k = \zeta_k(\omega)$ be the number of objects in the k-th generation $I_k, k = 0, 1, \ldots$. It can be verified that the sequence of random variables $\{\zeta_k, k = 0, 1, \ldots\}$ is a Galton–Watson branching process with generating function $f(s)$ (usually called the *embedded Galton–Watson process*). The essence of the proof is to verify the property

$$\mathrm{E}(s^{\zeta_{k+1}} | \zeta_1, \zeta_2, \ldots, \zeta_k) = [f(s)]^{\zeta_k} \tag{B.2}$$

which characterizes the Galton–Watson process. Equation (B.2) is a version of the forward pgf Eq. (3.5), conditional on ζ_k.

The embedded Galton–Watson process is helpful in proving that the probability of extinction for the Bellman–Harris process is subject to the same rules which govern the Markov versions. Let us note for example that if the embedded process becomes extinct for some ω, then the time-continuous process does too, since there is only a finite number of nonvoid generations $I_k(\omega)$ which may last for only a finite time. Thus $\lim_{k\to\infty} \zeta_k(\omega) = 0$ implies $\lim_{t\to\infty} Z(t,\omega) = 0$. The opposite is, in general, not true. An example can be a situation when all the objects in the k-th generation have lifelengths $\leq 2^{-k}$ and consequently $Z(t) = 0, t > 2$. The following result demonstrates that such occurrences have probability 0.

Theorem B.1 *Let A be the event $\{\zeta_n > 0$, for each $n\}$ and let B be the event $\{Z(t) > 0$, for each $t \geq 0\}$. If $P\{A\} > 0$, then $P\{B|A\} = 1$.*

Corollary B.1 *The probability of extinction, i.e., of the event* $\bar{B} \equiv \{Z(t) = 0,$ *for all sufficiently large* $t\}$*, is equal to the probability of the event* $\bar{A}$*, i.e., to the smallest nonnegative root* q *of the equation* $s = f(s)$*.*

B.2 Integral Equation

B.2.1 Decomposition into Subfamilies

If the initial object dies at or before time t, then the objects present at t are its children or their descendants. For a family history $\omega = (l, \nu; l_1, \nu_1; l_2, \nu_2; l_{11}, \nu_{11}; \dots)$ and each $i = 1, 2, \dots$, let us define $\omega_i = (l_i, \nu_i; l_{i1}, \nu_{i1}; l_{i2}, \nu_{i2}; l_{i11}, \nu_{i11}; \dots)$. The ω_i may be interpreted as the *family history of* $< i >$ *and its descendants*, although if $\nu < i$ then this family is not actually realized.

For the family history ω_i, let us define the random variables $Z_\iota(y, t, \omega_i)$, $Z(y, t, \omega_i)$, and $Z(t, \omega_i)$ in the way analogous to that in which, for ω, the rv's $Z_\iota(y, t, \omega)$, $Z(y, t, \omega)$, and $Z(t, \omega)$, were previously defined. Suppose that $l(\omega) \in [0, t]$ and $\nu(\omega) > 0$. It can be formally shown using the definitions above that

$$Z(t, \omega) = \sum_{i=1}^{\nu} Z(t - l, \omega_i). \tag{B.3}$$

In view of the fact that

$$I(\omega) = < 0 > \cup \bigcup_{i=1}^{\nu(\omega)} I(\omega_i),$$

the proof of (B.3) is reduced to careful "bookkeeping" of the indicator functions $Z_\iota(y, t - l, \omega_i)$ and $Z_{i\iota}(y, t, \omega)$.

B.2.2 Generating Functions

Let

$$F(s, t) = \sum_{r=0}^{\infty} \mathrm{P}\{Z(t) = r\} s^r. \tag{B.4}$$

Since the case $Z(t) = \infty$ has not yet been eliminated, it can be $F(1, t) < 1$. However, also in this case, the basic properties of the pgf's are verified. Let us note the alternative expression

$$F(s, t) = \mathrm{E}[s^{Z(t)}] \equiv \int_\Omega s^{Z(t,\omega)} d\mathrm{P}(\omega), \tag{B.5}$$

where $0^0 = 1$ and $s^\infty = 0$, even if $s = 1$.

Theorem B.2 *The generating function F satisfies the integral equation*

$$F(s,t) = s[1 - G(t)] + \int_{0-}^{t+} f[F(s,t-u)]dG(u), \tag{B.6}$$

where $t \geq 0$ and $s \in [0,1]$.

Proof Based on (B.5), let us write

$$F(s,t) = \int_\Omega s^{Z(t,\omega)} dP(\omega) = \int_{\{l>t\}} s^Z dP + \sum_{k=0}^{\infty} \int_{\{l\leq t, v=k\}} s^Z dP. \tag{B.7}$$

Since $Z(t,\omega) = 1$ if $l > t$, we have $\int_{\{l>t\}} s^Z dP = s \Pr\{l > t\} = s[1 - G(t)]$.

Let us consider Ω as a product space $\Omega' \times \Omega_1 \times \Omega_2 \times \dots$ of points $(l, v; \omega_1, \omega_2, \dots)$. Let P' be the probability measure on the pair (l, v) and let P_i be the probability measure on Ω_i. Now, it is possible to use (B.3). If l is fixed, then $Z(t-l, \omega_i)$ is a function on Ω_i, and hence if k is any positive integer we have

$$\int_{\{l\leq t, v=k\}} s^Z dP = \int_{\{l\leq t, v=k\}} dP'(l,v) \int_{\Omega_1} s^{Z(t-l,\omega_1)} dP_1 \dots \int_{\Omega_k} s^{Z(t-l,\omega_k)} dP_k.$$

Now each of the integrals $\int_{\Omega_i} s^{Z(t-l,\omega_i)} dP_i$ is equal to $F(s, t-l)$, since the probability measure $dP_i(\omega_i)$ is exactly the same as $dP(\omega)$. Hence the last equation is equal to $p_k \int_{0-}^{t+} [F(s,t-u)]^k dG(u)$. The same can be seen directly true if $k = 0$. Substitution into the right hand side of (B.7) yields the desired result.

B.2.3 Uniqueness of $F(s,t)$ and Finiteness of $Z(t)$

Theorem B.2 states that the pgf of $Z(t)$ satisfies Eq. (B.6), but it does not state that this solution is unique, nor that $\lim_{s\uparrow 1} F(s,t) = 1$ i.e., that $Z(t) < \infty$. We will outline here the arguments proving both these properties.

Regarding uniqueness, let us assume that there exists another pgf solution $\tilde{F}(s,t)$ of Eqn. (B.6). Then

$$|F(s,t) - \tilde{F}(s,t)| \leq \int_0^t |F(s,t-y) - \tilde{F}(s,t-y)| dG(y). \tag{B.8}$$

We see that since both F and $\tilde{F}$ are pgf's, $|F(s,t) - \tilde{F}(s,t)| \leq 1$. Substituting into the right hand side of (B.8) we obtain $|F - \tilde{F}| \leq G(t)$. Substituting this and repeating the estimate, we obtain that $|F - \tilde{F}| \leq G^{*i}(t)$ for any i. But $\lim_{i\to\infty} G^{*i}(t) = 0$ for any t (see Lemma 5.1), which yields $|F - \tilde{F}| = 0$.

Finiteness of $Z(t,\omega)$ may be obtained by estimating another random variable $\bar{Z}(t,\omega)$ equal to the *total number of objects* in family $I(\omega)$ that are born up to and

including time t (i.e., the *counting function of births*). Of course, $Z(t,\omega) \le \bar{Z}(t,\omega)$. We will consider the expected value of $\bar{Z}$. If it is finite, then $\bar{Z}$ is finite and so is Z (and consequently, $F(1-,t) = 1$.

For the argument, let us consider an object $< \iota > \neq < 0 >$, where $\iota = i_1 i_2 \dots i_k$. Let u_ι be a random variable that is 1 if $< \iota >$ is in the family $I(\omega)$ i.e., if it is ever born, and 0 otherwise, and let v_ι be a random variable that is 1 if $l + l_{i_1} + \dots + l_{i_1 i_2 \dots i_{k-1}} \le t$ and 0 otherwise. Then $< \iota >$ is born at or before t if and only if $u_\iota v_\iota = 1$, and

$$\bar{Z}(t) = 1 + \sum_{k=1}^{\infty} \sum_{i_1 i_2 \dots i_k = 1}^{\infty} u_{i_1 i_2 \dots i_k} v_{i_1 i_2 \dots i_k}.$$

Expected value $\mathrm{E}(v_\iota)$ is equal to $G^{*k}(t)$. The rv u_ι is the indicator function of the event that object $< \iota >$ is ever born and therefore its expectation is equal to the probability of this event, i.e., to

$$\begin{aligned} \mathrm{E}(u_\iota) &= \mathrm{P}\{v \ge i_1, v_{i_1} \ge i_2, \dots, v_{i_1 \dots i_{k-1}} \ge i_k\} \\ &= \mathrm{P}\{v \ge i_1\} \mathrm{P}\{v_{i_1} \ge i_2\} \dots \mathrm{P}\{v_{i_1 \dots i_{k-1}} \ge i_k\}. \end{aligned}$$

The u_ι's and v_ι's are independent, so that

$$\begin{aligned} \mathrm{E}[\bar{Z}(t)] &= 1 + \sum_{k=1}^{\infty} G^{*k}(t) \sum_{i_1} \mathrm{P}\{v \ge i_1\} \sum_{i_2} \mathrm{P}\{v_{i_1} \ge i_2\} \dots \sum_{i_k} \mathrm{P}\{v_{i_1 \dots i_{k-1}} \ge i_k\} \\ &= 1 + \sum_{k=1}^{\infty} G^{*k}(t) [f'(1-)]^k. \end{aligned}$$

Lemma 5.1 states that this sum is $< \infty$ for all t and so $\mathrm{E}[\bar{Z}(t)] < \infty$.

Appendix C
General Processes

C.1 Introduction to the Jagers-Crump-Mode Process

This section is a useful reference but it can be omitted at first reading. Its aim is to introduce the reader in an informal way to the basics of the general branching processes. In most part, the book is concerned with less general processes and therefore the subject can be postponed to a later reading. However, there are issues that are best expressed when phrased in terms of general processes. An example is an application of a general process to cell populations in Sect. C.2.

C.1.1 Definition of the General Branching Process

A basic source concerning general branching processes is the book by Jagers (1975). Our account is also based on Taïb (1992).

Individuals

We consider development in time of a population started by a single individual. The individuals can be considered elements of the set

$$I = \bigcup_{n=0}^{\infty} N^n,$$

called the Ulam–Harris space, where $N = \{1, 2, \dots\}$ and $N^0 = \{0\}$. Individual 0 is the ancestor of the population. Each element of N^n is of the form $x = (x_1, \dots, x_n)$. The meaning of this notation is that the individual belongs to the n-th generation and is the x_n-th progeny of the x_{n-1}-st progeny, $\dots$, of the x_1-th progeny of the ancestor. This description is redundant, as not all these individuals will come to existence in a given realization of the process. Each of the individuals evolves in a space Ω, which

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology*,
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

is large enough to allow for all possible life-spans and progeny-bearing processes of this individual. An element $\omega \in \Omega$, is this individual's life. The probability measure on a σ-algebra $\mathcal{F}$ of Ω is called Q.

Lives

For each individual $\tau(\omega, k),\ k = 1, 2, \ldots$ denote successive ages at childbearing. In particular, $\tau(\omega, k)$ is the age at which the individual has its k-th progeny. These ages are organized as epochs of a point process, a random collection of time moments or equivalently a random collection of nonnegative-integer valued measures, denoted ξ. Mathematically,

$$\xi(\omega, [0, t]) = \xi(t) = \#\{k : \tau(\omega, k) \leq t\},$$

is the counting function of births, i.e., the number of progeny begotten before or at the age of t. In addition, λ, the duration of life ω of an individual, is a random variable $\lambda : \Omega \to R^+$.

The time evolution of the individuals is governed by the connections between their times of births. Let σ_x denote the moment of birth of individual x ($\sigma_0 = 0$, for the ancestor). Then, if we denote by xk the individual being the k-th progeny of x, we set

$$\sigma_{xk} = \sigma_x + \tau_x(k).$$

In this latter expression, the argument ω is dropped, as it will be frequently done.

Construction of the Process

If the space Ω is a Polish space (i.e., it is metric, complete and separable), then the σ-algebra $\mathcal{F}$ can be selected as the class of Borel sets of Ω. The triplet $(\Omega, \mathcal{F}, Q)$ is the probability space of a single individual. If we assume that the lives of individuals are independent, then the space of the process can be constructed as a product space of the form $(\Omega^I, \mathcal{F}^I, Q^I)$, where I is the collection of all individuals. From now on, we will write P instead of Q^I and ω instead of $\{\omega_x, x \in I\}$.

The model presented can be specialized to include the classical branching processes, by assuming that all $\tau(\omega, k),\ k = 1, 2, \ldots$ are concentrated at $\lambda(\omega)$, i.e., all progeny are born at the same time. Then, if $\lambda(\omega) = 1$, we obtain the Galton-Watson process. If $\lambda(\omega)$ is a nonnegative rv, we obtain the Bellman-Harris process, etc.

C.1.2 Random Characteristics and Basic Decomposition

The method of random characteristics makes it possible to account for individuals existing in the process, individuals being born during a given time interval, individuals with ages from a given interval, individuals with a given number of progeny,

etc. The random characteristic is a random function $\chi_x(a)$ defined on an individual's life. It defines the contribution, of a desired type, of individual x, from its birth until it reaches age a. The summary contribution of all individuals at a given time t, is equal to

$$Z_t^{\chi} = \sum_{x \in I} \chi_x(t - \sigma_x).$$

Z_t^{χ} is called the process counted by random characteristic $\chi_x(a)$. For example, if

$$\chi_x(a) = \begin{cases} 1, & \text{if } a \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

then Z_t^{χ} counts all individuals born until time t. If

$$\chi_x(a) = \begin{cases} 1, & \text{if } a \in [0, \lambda_x), \\ 0, & \text{otherwise,} \end{cases} \tag{C.1}$$

then Z_t^{χ} counts all individuals alive at time t. If

$$\chi_x(a) = \begin{cases} 1, & \text{if } a \in [0, \lambda_x) \cap [\tau_x(k), \infty), \\ 0, & \text{otherwise,} \end{cases}$$

then Z_t^{χ} counts all individuals alive at time t, with at least k progeny born before t.

For the process counted by random characteristics, it possible to write a backward decomposition, analogous to (1.1)

$$Z_t^{\chi} = \chi_0(t) + \sum_{i=1}^{X} Z_{t-\tau_0(i)}^{\chi\,(i)},$$

where X is the number of progeny effectively begotten by the ancestor and superscript (i) denotes the i-th iid copy of the process.

C.1.3 Expectations, Malthusian Parameter and Exponential Growth

Reproductive measure is the expectation of the point process of progeny births

$$\mu(A) = \mathrm{E}[\xi(\omega, A)].$$

It is characterized by the reproductive counting function $\mu(a) = \mu([0, a])$. The expectation of the process, $m_t = \mathrm{E}(Z_t^{\chi})$ counted by characteristic $\chi(a)$ with expectation

$g(a) = \mathrm{E}[\chi(a)]$ can be represented by the expression

$$m_t = \sum_{n=0}^{\infty} \int_0^t g(t-u) d\mu^{*n}(u) = \int_0^t g(t-u) d\nu(u),$$

where $\nu(u) = \sum_{n=0}^{\infty} \mu^{*n}(u)$. The n-th convolution power of the reproductive measure, μ^{*n}, counts the expected number of progeny born to the n-the generation individuals in the process. Then, each of μ^{*n} has to be convolved with the expectation of the random characteristic, to account for proper bookkeeping, and the result summed over all generations of the process. Under mild conditions (e.g., no concentration of births at age 0 and expected total progeny of an individual finite), this sum is finite. Expectation m_t satisfies a renewal-type integral equation

$$m_t = \int_0^t m_{t-u} d\mu(u) + g(t). \tag{C.2}$$

A major role in the theory is played by the Malthusian parameter, which determines (if it exists) the asymptotic rate of growth of m_t. The Malthusian parameter is the real solution of the equation

$$\hat{\mu}(\alpha) \equiv \int_0^{\infty} e^{-\alpha u} d\mu(u) = 1.$$

This solution, if it exists, is unique. In what follows, we will limit ourselves to the supercritical case that is when $\mu([0,\infty)) > 1$ [see the classification (1.5)]. In this case the Malthusian parameter exists and is positive. The renewal theorem demonstrates, in the same way as it was explained in Sect. 5.2 for the Bellman–Harris process, that m_t behaves asymptotically like $e^{\alpha t}$,

$$e^{-\alpha t} m_t \longrightarrow \underbrace{\frac{\int_0^{\infty} g(u) e^{-\alpha u} du}{\int_0^{\infty} u e^{-\alpha u} d\mu(u)}}_{\beta} \equiv c(\chi), \quad \text{as } t \to \infty. \tag{C.3}$$

If we assume that all progeny are born at the same time τ in the life of the individual, so that $\mu(u) = mG(u)$, where m is the mean count of progeny and $G(\,\cdot\,)$ is the cumulative distribution of τ, and that this is exactly the moment of individual's death, i.e., that $\lambda = \tau$, we obtain the Bellman–Harris process of Chap. 5. If we wish to account for individuals alive at time t, then we use the random characteristics of the form $\chi_x(a) = 1$, if $a \in [0, \tau)$, and $\chi_x(a) = 0$, otherwise, as in Eqn. (C.1). This means that $g(u) = 1 - G(u)$. Substituting into (C.3), we obtain the expression derived for the Bellman–Harris process (5.13)

Without getting into more detail, we state that in the supercritical case, the entire process counted by a random characteristic behaves very much the same way as its expectation. Indeed, there exists a random variable W, with $\mathrm{E}(W) = 1$, such that

$$Z_t^{\chi} e^{-\alpha t} \longrightarrow c(\chi) W,$$

as $t \to \infty$, with probability 1.

C.1.4 Abstract Type Spaces and Composition of the Process

Let us suppose that each newborn individual is endowed, at birth, with a type selected from a measurable space $(\Gamma, \mathcal{G})$, where $\mathcal{G}$ is a σ-algebra of subsets of Γ. In other words, there exist measurable mappings $\rho(j) : \Omega \to \Gamma$, which determine the types of newborn individuals. The point process ξ, which describes reproduction, is now defined by

$$\xi(A \times B) = \#\{i \in N; \rho(i) \in A, \tau(i) \in B\}.$$

Intuitively, $\xi(A \times B)$ is the number of progeny of an individual, born in time set B, with types in set A. The population of individuals can be defined on $(\Gamma \times \Omega^I)$, where Γ describes the type of the ancestor. The theorem of Ionesco-Tulcea enables to construct a unique probability measure P_γ on $(\Gamma \times \Omega^I, \mathcal{G} \times \mathcal{A}^I)$ for the process with a type γ ancestor. Similarly as before, a major role is played by the reproduction kernel $\mu(\gamma, A \times B) = \mathrm{E}_\gamma[\xi(A \times B)]$. For each real λ, we define

$$\mu_\lambda(\gamma, d\gamma' \times du) = e^{-\lambda u}\mu(\gamma, d\gamma' \times du)$$

and

$$\hat{\mu}_\lambda(\gamma, d\gamma') = \int_0^\infty \mu_\lambda(\gamma, d\gamma' \times du).$$

The Malthusian parameter α is selected so that the kernel $\hat{\mu}_\alpha(\gamma, d\gamma')$ has a Perron-Frobenius eigenvalue equal to 1 (assuming this latter exists). The Perron-Frobenius eigenvalue is the real eigenvalue strictly dominating absolute values of all remaining eigenvalues. If we set $\nu_\alpha(\gamma, d\gamma' \times du) = \sum_{n\geq 0} \mu_\alpha^n(\gamma, d\gamma' \times du)$, where $\mu_\alpha^n(\gamma, d\gamma' \times du)$ is the n-fold convolution of measure $\mu_\alpha(\gamma, d\gamma' \times du)$ with respect to elements $d\gamma' \times du$, we can write

$$\mathrm{E}_\gamma[e^{-\alpha t} Z_t^\chi] = \int_{\Gamma \times R_+} \mathrm{E}_\gamma[e^{-\alpha(t-u)}\chi(t-u)]\nu_\alpha(\gamma, d\gamma' \times du).$$

So, we see that $\mathrm{E}_\gamma[Z_t^\chi]$ is of the form $R * g(\gamma, t)$, where $R = \nu_\alpha$ and

$$g(\gamma, t) = \mathrm{E}_\gamma[e^{-\alpha t}\chi(t)].$$

Asymptotic behavior of the expectation of the process and of the process itself in the supercritical case ($\alpha > 0$) depends on the conservativeness of the kernel $\hat{\mu}_\alpha(\gamma, d\gamma')$. For countably generated $\mathcal{G}$ the kernel is conservative if its potential $\hat{\nu}_\alpha(\gamma, d\gamma') = \sum_{n\geq 0} \hat{\mu}_\alpha^n(\gamma, d\gamma')$ has the property that there exists a σ-finite measure m on $(\Gamma, \mathcal{G})$ such that

$$m(A) > 0 \implies \hat{\nu}_\alpha(\gamma, A) = \infty \tag{C.4}$$

for all $\gamma \in \Gamma$. This property is a generalization of positive regularity of matrices.

If the kernel $\hat{\mu}_\alpha$ is conservative, there exists an eigenfunction h, satisfying

$$\begin{aligned} h(\gamma) &= \int_{R_+}\int_\Gamma e^{-\alpha u}h(\gamma')\mu(\gamma, d\gamma' \times du) \\ &= \int_\Gamma h(\gamma')\mu_\alpha(\gamma, d\gamma'). \end{aligned} \tag{C.5}$$

So, $e^{-\alpha u}\frac{h(\gamma')}{h(\gamma)}\mu(\gamma, d\gamma' \times du)$ has total mass on $\Gamma \times R_+$ equal to 1 and it is a probability measure. $h(\gamma)$ is the reproductive value of individuals of type γ. It indicates the relative long-term contribution of individuals of this type to the population.

If the kernel $\hat{\mu}_\alpha$ is conservative, there also exists a probability measure π, which satisfies

$$\pi(d\gamma') = \int_\Gamma \hat{\mu}_\alpha(\gamma, d\gamma')\pi(d\gamma). \tag{C.6}$$

This equation can also be written in the following manner:

$$h(\gamma')\pi(d\gamma') = \int_\Gamma \frac{h(\gamma')}{h(\gamma)}\hat{\mu}_\alpha(\gamma, d\gamma')h(\gamma)\pi(d\gamma)$$

if $\inf h(\gamma) > 0$. We can then normalize the equation so that we obtain $\int_\Gamma h(\gamma)\pi(d\gamma) = 1$. Measure π defined above can be interpreted as a stable distribution of the types of the newborns. Consequently, an individual drawn at random from a very old population is of a random type decided by π, independently of the initial conditions.

Another interesting expression:

$$\beta = \int_\Gamma \int_\Gamma \int_{R_+} te^{-\alpha t}h(\gamma')\mu(\gamma, d\gamma' \times dt)\pi(d\gamma)$$

can be considered the expected age at reproduction.

Similarly as in the single-type case, in the supercritical case ($\alpha > 0$) a generalization of the key renewal theorem makes it possible to calculate the limit of $\mathrm{E}[e^{-\alpha t}Z_t^\chi]$. We will denote $\mathrm{E}_\pi(X) = \int_\Gamma \mathrm{E}[X]\pi(d\gamma)$, the expectation in the process with the type of ancestor being randomly drawn according to measure π. Then we have

$$\mathrm{E}[e^{-\alpha t}Z_t^\chi] \longrightarrow \frac{\mathrm{E}[\hat{\chi}(\alpha)]}{\alpha\beta}h(\gamma),$$

as $t \to \infty$, for all γ except sets of π-measure 0. The process behaves in the supercritical case very much like its expectation.

The multitype formulation provides a great generality and was used in applications, particularly concerning evolution theory (Taïb, 1992).

C.2 Application: Alexandersson's Cell Population Model Using a General Branching Process

An elegant example of modeling using general processes and counting characteristics (Sect. C.1) is a part of Alexandersson's (1999) thesis. This application demonstrates how a branching process approach complements existing deterministic approaches, while the construction of the process is very straightforward.

C.2.1 The Model

Let us consider a cell population, where each cell inherits a type at birth, grows during a stochastic time span, and when its cell cycle is completed it divides into two not necessarily equal daughter cells. The type of the individual is the birth size of the cell expressed as mass, volume, DNA-content etc. Since cells have only two progeny, the Ulam-Harris space of all possible cells reduces to

$$I = \bigcup_{n=0}^{\infty} \{1,2\}^n,$$

where $\{1,2\}^0 = \{0\}$.

The type space is an interval $S = (0, M]$ of the real line, where $M < \infty$ is the largest possible birth size of a cell, and $\mathcal{S}$ is the Borel-σ-algebra on S. A cell with birth size $r \in S$ chooses a life ω from $(\Omega, \mathcal{A})$ using $\mathrm{P}(r, \cdot)$, the life law of cells of type r.

We construct the population space $(S \times \Omega^{\mathcal{I}}, \mathcal{S} \times \mathcal{A}^{\mathcal{I}})$ as in Sect. C.1. Under the assumption that the daughter processes of different cells are conditionally independent, there exists a unique probability measure P_r on the entire population process, where $r \in S$ is the type of the ancestor.

The size of a cell with initial size r increases with time according to a deterministic growth function g. We let $m(r,t)$ denote the size of an r -type cell at age t. The functions m and g are related by the initial value problem

$$\frac{dm}{dt} = g(m), \quad m(r,0) = r.$$

The cell grows and, after division, the daughter cells do not necessarily have the same size (type) at birth. Note that we do not allow cell death in this model, so our branching population is supercritical. Let λ denote the age of the cell at division (the cell cycle time) and let the distribution of λ be defined by its hazard rate function $b(s)$, $s \in (0, 2M]$, i.e., $\mathrm{P}[\lambda > s] = \exp[-\int_0^s b(u)du]$

A cell of type r divides into fractions δ and $1-\delta$, where δ is a random variable on $(0,1)$ with density function $f_\delta(m,p)$, $p_1 \le p \le p_2$, where $p_1 = 1 - p_2 \in (0,1)$ depends on $m = m(r,\lambda)$, the cell size at division. We will assume that f_δ is unimodal

and that δ is symmetrically distributed around $1/2$, i.e., for all $r \in S$, $f_\delta(m,p) = f_\delta(m,1-p)$, and $\mathrm{E}_r[\delta] = 1/2$.

Let $T(x) = \int_0^x \frac{1}{g(y)} dy$, $x \in S$. To see how to interpret this function, consider

$$T(x) - T(r) = \int_r^x \frac{1}{g(y)} dy. \tag{C.7}$$

Making a change of variable $y = m(r,t)$ yields $dy = dm(r,t) = g(m(r,t))dt$ and (C.7) becomes

$$\int_0^u \frac{g(m(r,t))}{g(m(r,t))} dt = \int_0^u dt = u,$$

where u is the time it takes for a cell to grow from size r to size x. Consequently, $T(x) - T(r)$ is precisely this time. Since $T(m(r,t)) - T(r) = t$, $m(r,t) = T^{-1}(T(r) + t)$. Further let $C(x) = \int_0^x \frac{b(y)}{g(y)} dy$ and $Q(x) = b(x)/[xg(x)]$, and assume that each cell has to divide before it reaches size $2M$, i.e., b is such that for $\epsilon > 0$

$$\int_0^{2M} \frac{b(y)}{g(y)} dy = \infty \quad \text{and} \quad \int_0^{2M-\epsilon} \frac{b(y)}{g(y)} dy < \infty.$$

The reproduction kernel $\mu(r, ds \times dt)$, which is the expected number of children with birth sizes in ds to a cell of type r with age in dt, takes the form

$$\begin{aligned}
\mu(r, ds \times dt) &= \mathrm{E}_r[\xi(ds \times dt)] \\
&= \mathrm{E}_r[\mathbf{1}(\lambda \in dt)(\mathbf{1}(\delta m(r,\lambda) \in ds) + \mathbf{1}((1-\delta)m(r,\lambda) \in ds))] \\
&= 2 \int_0^\infty \mathbf{1}(u \in dt) \int_0^1 \mathbf{1}(pm(r,u) \in ds) f_\delta(m(r,u), p) dp \\
&\quad b(m(r,u)) e^{-\int_0^u b(m(r,v))dv} du,
\end{aligned}$$

where the factor 2 comes from the fact that δ and $(1-\delta)$ are identically distributed.

The inner integral is zero everywhere except when $p = s/m(r,u)$ and $dp = ds/m(r,u)$ so we have

$$\begin{aligned}
\mu(r, ds \times dt) = 2 \int_0^\infty \mathbf{1}(u \in dt) f_\delta(m(r,u), s/m(r,u)) \frac{b(m(r,u))}{m(r,u)} \\
e^{-\int_0^u b(m(r,v))dv} du\, ds.
\end{aligned} \tag{C.8}$$

Making a change of variable in the same manner as above, with $x = m(r,u)$, we get that $du = \frac{dx}{g(x)}$ and the kernel becomes

$$\mu(r, ds \times dt) = 2 \int_r^{2M} \mathbf{1}(T(x) - T(r) \in dt) f_\delta(x, s/x) Q(x) e^{-(C(x)-C(r))} dx ds.$$

C.2.2 Existence of the Stable Birth Size Distribution

If the Malthusian parameter α exists such that $\hat{\mu}_\alpha$ is conservative, then the Perron–Frobenius Theorem gives the existence of a function h (see (C.5)) and a measure π (see (C.6)). By requiring strong or positive α-recurrence (Jagers and Nerman 1996) and $\inf h > 0$ we can norm to

$$\int_S h(s)\pi(ds) = 1, \quad \int_S \pi(ds) = 1.$$

The measure π is then called the stable birth type distribution. Hence we want to prove the existence of the Malthusian parameter, i.e., prove the existence of a number $\alpha > 0$ such that the Perron-root $\rho(\hat{\mu}_\alpha) = 1$, where

$$\hat{\mu}_\alpha(r, A) = \int_{\mathrm{R}_+} e^{-\alpha t}\mu(r, A \times dt)$$

and also that $\hat{\mu}$ is conservative.

Theorem C.1 *Under the assumptions stated in Sect. C.2.1 on the reproduction kernel μ, the Malthusian parameter α exists and the kernel $\hat{\mu}_\alpha$ is conservative.*

C.2.3 Asymptotics of the Cell Model

We discuss the asymptotics of our cell model. When looking at a population one can either consider all cells alive at the moment, or all cells born into the population up till now. Even if it seems more natural to look at all cells alive, it is mathematically more convenient to consider all born. In this chapter we will concentrate on all born cells, but we will also show that all the results presented can easily be obtained for all cells alive as well. When calculating the asymptotics of our model, we construct random characteristics used to count the population with respect to some property. An alternative way, described in (Jagers and Nerman 1996), is to sample an individual at random in an already stabilized population, and consider the population with time centered around this individual. The individual sampled at random is called ego.

The α-curve is the graph of the function $\alpha(a)$ describing the proportion of cells still undivided at age a. An alternative interpretation is that $\alpha(a)$ is the probability that the age at division of a cell sampled at random, ego, is larger than a. In order to find an expression for $\alpha(a)$ we define a random characteristic χ (*c.f.* Chap. C.1) such that z_t^χ counts the number of cells born up to time t with respect to χ. Then, if y_t denotes the number of all cells born up to time t, we can use the result that under suitable conditions

$$\frac{z_t^\chi}{y_t} \to \mathrm{E}_\pi[\hat{\chi}(\alpha)] \text{ as } t \to \infty$$

in probability (on the set of non-extinction), where $\mathrm{E}_\pi[X] = \int_S \mathrm{E}_s[X]\pi(ds)$, $\hat{\chi}(\alpha) = \int_{\mathrm{R}_+} \alpha e^{-\alpha t}\chi(t)dt$ and π is the stable birth type distribution.

The random characteristic that gives score one for each cell x born up to time t and with life length λ_x longer than a can be written as

$$\chi_x(t) = \mathbf{1}_{\mathrm{R}_+}(t - \tau_x)\mathbf{1}(\lambda_x > a)$$

where τ_x is the birth time for cell x. Making a change of variable $u = t - \tau_x$ gives

$$\chi(u) = \mathbf{1}_{\mathrm{R}_+}(u)\mathbf{1}(\lambda > a).$$

This yields

$$\begin{aligned}
\alpha(a) = \mathrm{E}_\pi[\hat{\chi}(\alpha)] &= \int_S \mathrm{E}_r[\hat{\chi}(\alpha)]\pi(dr) \\
&= \int_S \mathrm{E}_r[\int_{\mathrm{R}_+} \alpha e^{-\alpha u}\chi(u)du]\pi(dr) \\
&= \int_S \int_{\mathrm{R}_+} \alpha e^{-\alpha u}du\,\mathrm{E}_r[\mathbf{1}(\lambda > a)]\pi(dr) \\
&= \int_S \mathrm{P}_r(\lambda > a)\pi(dr) \\
&= \int_S e^{-\int_0^a b(m(r,v))dv}\pi(ds).
\end{aligned}$$

The β-curves are used to describe the proportions of sister cells, cousin cells etc., with life lengths that differ by more than a time units. The β_1-curve describes this proportion for sister cells, β_2 for cousin cells and so on. Alexandersson's (1999) thesis includes further asymptotic results for the β-curves and numerical computations for the model we outlined. Furthermore, it also deals with a much more complicated example of cell proliferation, which we consider, using different methods, in Sect. 7.8.2.

Glossary

Biological Glossary for Mathematicians

Cross-references to other glossary terms are *italicized*.

Acquired immune deficiency syndrome (AIDS) A disease associated with the *Human Immunodeficiency Virus*. It is characterized by the viral inactivation of some of the host's cytotoxic T cells, one of the important components of the immune system. This results in a decrease in ability of the host to immunologically respond to foreign entities such as viruses, bacteria, fungi, and cancer cells.

AIDS See *Acquired Immune Deficiency*

***Alu* elements** A kind of *repeat DNA* sequence of about 300 base pairs dispersed in many locations throughout the human *genome*. *Alu* elements do not code for proteins.

Amino acids The twenty different basic units of proteins.

Amplification (Gene Amplification) The increase in the number of copies of a *gene*. May result from errors in DNA *replication* or *recombination*.

Antibody A *protein* produced by the immune system in response to a foreign molecule (*antigen*) that interacts specifically with the foreign molecule.

Antigen A molecule that induces an *antibody*.

Apoptosis A kind of programmed cell death in which the *cell* contents are conservatively degraded. The *DNA* is cut into discrete large pieces, and the cytoplasm contents are redistributed into several smaller membrane bound packages. These may be engulfed by other cells, the molecules processed, and reused.

Bacteria *Cells* of a lower form of life without a *nuclear* membrane.

Base pair Usually used as a unit of length of a DNA strand, spanning one pair of complementary nucleotides.

Cancer A population of *cells* that continue to divide and survive under conditions in which normal cells would stop dividing or die. The cancer cell population usually is thought to be initiated from a single cell (clonal origin). As the progeny of the single cell multiply they accumulate *mutations* and acquire new characteristics

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

(tumor progression). They may invade adjacent tissues, and travel to distant sites to form secondary tumors (metastases).

Cancer stem cell A *stem cell* within a tumor that has the ability to divide and produce more cells and the ability to divide and produce differentiated cells. The hierarchal model of cancer stem cells describes only a subset of cells within the tumor as having a high probability of acting as stem cells. In contrast the stochastic model of cancer stem cells describes all cells in the tumor as having an equal probability of acting as stem cells. A different use of the same term, cancer stem cell, refers to the cell that initiates a *colony* of cancer cells that is the tumor.

Cell The basic unit of life. Cells of higher forms of life have an outer membrane surrounding the cytoplasm and the *nucleus*. In the cytoplasm there are *proteins* (enzymes) that carry out biochemical functions, machinery (ribosomes) for making proteins, and compartments (organelles) such as *mitochondria*. Higher forms of life, such as mammalian cells, which have a membrane surrounding their nucleus, are referred to as eukaryotes. Lower forms of life, such as *bacteria*, which do not have a membrane surrounding their nucleus, are referred to as prokaryotes.

Cell cycle The stages of *cell* growth and division. Includes the following stages (phases): division of one cell to produce two cells (cytokinesis), a gap of time (G_1 phase) between cytokinesis and the initiation of *DNA* synthesis (S phase), a gap of time (G_2 phase) between the end of DNA synthesis and the formation of visible chromosomes, and *mitosis* (M phase). In mitosis the duplicated *chromosomes* (chromatids) containing replicated *DNA* are partitioned to new *cells* at cell division. The time between one cell division and another is referred to as the cell lifetime.

Centromere A part of the *chromosome* required for proper movement of the daughter chromosomes (chromatids) to daughter *cells*. A piece of *DNA* that is not part of a chromosome and does not contain a centromere DNA sequence is referred to as an acentric extrachromosomal element or double minute chromosome. Such acentric extrachromosomal elements do not segregate properly into daughter cells.

Chemotherapy The treatment of *cancer* cells with chemicals that kill them. In combination drug therapy two or more chemicals with different modes of action are used to increase the efficiency of killing cancer cells.

Chromosome The linear structure containing *DNA* and *protein* that is visible under a microscope at *mitosis*. Chromosomes contain DNA sequences (*genes*) that code for proteins, and DNA sequences that do not code for proteins. Among the non-coding DNA sequences there are *centromeres* necessary for the separation of daughter chromosomes (chromatids) during mitosis, and *telomeres* which function to maintain the integrity of the ends of chromosomes.

Chronic myeloid leukemia (CML) A kind of *leukemia* that is the result of a *chromosome* defect in a bone marrow *stem cell*, resulting in a *colony* of abnormal cells in the blood. After several years in the chronic phase most patients undergo a sudden change, blast crisis, to an acute phase in which myeloid and/or lymphoid cells are overproduced.

Colony A population of *cells* that are the progeny of a single cell.

DNA Deoxyribonuleic acid. The genetic material. A long double helix with a structure similar to a twisted ladder. The backbones of the ladder are strands composed

of alternating sugar (deoxyribose) and phosphate groups. The rungs of the ladder are pairs of nucleotide subunits. The nucleotide subunits are abbreviated *A, T, G, C*. *A* is paired with *T*, and *G* is paired with *C*. The genetic information in DNA is stored in the sequence of nucleotides. The information is transcribed into complementary copies of a sequence of nucleotides in messenger RNA and is then translated into a sequence of *amino acids* in *protein*. During DNA replication the two strands of a double helix separate and each acts as a template to synthesize a new complementary strand. Each of the two double helices (one new strand and one old strand) is contained in each one of a pair of sister chromatids (the daughters of *chromosomes*). The sister chromatids segregate into daughter cells at mitosis.

Drug resistance The continued survival of *cells* in the presence of chemicals (drugs) intended to kill them. Resistance to two or more drugs is referred to as double resistance or cross resistance.

Eve The hypothetical common human female ancestor of all extant humans. Suggested by some common genetic features of individuals in current human populations.

Flow cytometry A method for the analysis of the distribution of the amount of a molecule (such as *DNA* or *protein*) in a population of *cells*. Cells are stained and pumped through a thin tube between a light source and a detector. Measurements of the amount of DNA per cell are used to indicate the number of cells in each phase of the cell cycle. Measurements of the amount of a specific protein per cell are used to indicate overproduction of the protein as a result of, for instance, *gene amplification*.

Fluctuation analysis Also, Luria and Delbrück fluctuation analysis. A method to determine *mutation* rates of bacteria or mammalian cells. Parallel cultures of cells are grown for a number of generations, and then the number of mutants in each culture, the average number of mutants per culture, and the number of cultures containing no mutants are determined. This information can be used to calculate the mutation rate, i.e. the number of new mutations per cell per generation. In contrast to mutation rate, the mutation frequency is the number of mutant cells in a culture at one time.

Gene A sequence of bases in *DNA* that codes for a *protein* and influences the inherited characteristics of a *cell* or organism.

Genome All of the sequence of bases of *DNA* or *RNA* of a *cell*, or of a *virus*. The information in the genome functions in inheritance and to determine the characteristics of the organisms. Some DNA sequences may code for proteins that influence the inherited characteristics of a *cell*, or of a *virus*. Other DNA sequences may have a regulatory or structural function. The genome of cells is DNA, the genome of some viruses is DNA, and the genome of other viruses is RNA.

Heterogeneity (Tumor heterogeneity) Populations of *cancer* cells that contain subpopulations with different characteristics, such as relative resistance to drugs.

HIV See *Human immunodeficiency virus*

Human Immunodeficiency Virus (HIV) The RNA containing *retrovirus* that is the cause of *Acquired immune deficiency (AIDS)* disease.

Leukemia A kind of *cancer* of the blood forming cells.

Mitosis The stage of the cell cycle of somatic (body) *cells* in which replicated *chromosomes* (chromatids) are separated into daughter cells. The result of mitosis is two daughter cells that have identical sets of *genes*. Daughter cells may be slightly different in size as a result of asymmetric division of the cytoplasm at cell division.

Meiosis The formation of gametes (sex *cells*) by two successive cell divisions and only one round of *DNA* synthesis. This results in the segregation of non-identical forms of genes (alleles) into different gametes. The gametes are haploid, containing half as much DNA as diploid body cells.

Mitochondria Organelles in the cytoplasm of *cells* of higher organisms needed for generating energy. Mitochondria contain *DNA*. They are inherited only from the mother, hence the term maternal inheritance.

Molecular clock hypothesis The assumption that *mutations* in a *gene* occur randomly and at an approximately equal rate over long time intervals during evolution.

Mutant An organism or *cell* that has a different inherited characteristic than the remainder of the cells in a population. Usually the result of a change in *DNA* sequence.

Mutation A change in *DNA* sequence. Usually detected by a sudden and inherited change in an observed characteristic (*phenotype*) of a *cell* or of an organism. However, a mutation may be detected directly by determining a change in the DNA sequence, even though there is no visible characteristic change in the cell or organism. The progeny of the mutant may revert to the previous phenotype, in which case the new mutation is referred to as a reverse mutation or back mutation. A phenotype resulting from a series of two mutations is referred to as a two stage mutation. The rate of mutation may be determined by *fluctuation analysis*.

Nucleus The part of a *cell* containing *DNA*. The part of the cell outside of the nucleus is referred to as the cytoplasm.

Oncogene A *gene* (*DNA* sequence) associated with *cancer*. An oncogene can be detected and mapped by its pattern of inheritance in cancer prone families. A piece of DNA containing an oncogene can be detected by the ability of the DNA to induce cancer-like changes when transferred into *cells* growing in culture.

Organelle A part of a *cell* which carries out a specialized function. An example is a mitochondrion (plural: mitochondria). A mitochondrion is a *DNA* containing, membrane enclosed, structure located in the cytoplasm. It functions to produce high energy molecules for cell metabolism. During cell division, mitochondria may, or may not, be distributed to daughter cells in equal numbers.

PCR See *Polymerase chain reaction*.

Phenotype The visible characteristics of a *cell* or organism. As opposed to genotype, the genetic information of a cell.

Plasmid In bacteria, a circular piece of *DNA* that is separate from the major (chromosomal) piece of DNA. Plasmids replicate and segregate at *cell* division independently of the chromosomal DNA. Each bacterial cell may contain multiple numbers of plasmids which may be randomly distributed at cell division.

Polymerase chain reaction An experimental procedure for obtaining a large number of copies of a piece of *DNA*. The procedure employs short pieces of DNA complementary to the ends of the desired sequence, and the enzyme DNA polymerase to exponentially increase the number of copies of the desired DNA sequence.

Population biology The study of groups of people in time and space. Variables may include genes, proteins, physical characteristics, ethnicity, etc.

Protein A polymer molecule consisting of monomer subunits of *amino acids*. The linear sequence of amino acids in a protein is determined by the corresponding sequence of nucleotides in *DNA* (*gene*). Some proteins (enzymes) function to encourage chemical reactions, while other proteins have a structural function.

Quiescence A phase when *cells* are pausing rather than actively progressing through the *cell cycle*. Most cells of higher organisms are quiescent rather than actively dividing. They pause before the initiation of *DNA* synthesis.

Recombination The formation of new combinations of *genes* by the exchange of genetic information between *chromosomes*.

Repeat DNA Sequences of *DNA* nucleotides that are tandemly iterated. In some diseases the number of repeats may vary between individuals, and the number may change from parents to progeny.

Replication The duplication of *DNA*. Two strands of DNA separate, like a zipper, at a moving replication fork. Each strand acts as a template to code for a complementary sequence of nucleotides in a new strand. The result is two new pieces of DNA, each double stranded, and each piece containing one new strand and one old strand. This is referred to as semiconservative replication. Errors may occur during DNA replication, slippage at the replication fork or redundant replication forks, resulting in sequences that are added or deleted (*amplification* or deamplification).

Retrovirus See *Virus*

RNA Ribonucleic acid A molecule similar to *DNA*, but with a different sugar (ribose rather than deoxyribose), one different nucleotide (*U* rather than *T*), and mostly single stranded (rather than double stranded). There are several kinds of RNA. One of these, messenger RNA (mRNA), is transcribed as a complementary copy of the sequence of nucleotides in DNA and functions to determine the sequence of *amino acids* in *protein*. RNA can also function as the *genome* of some *viruses*, called *retroviruses*.

Segregation The separation of different forms of *genes* (alleles) into sex *cells* (gametes) by the separation of chromatids (daughter *chromosomes*) at the second cell division of *meiosis*. Also, the separation of chromatids to daughter cells during *mitosis*.

Senescence The inability of some normal *cell* populations to continue to divide indefinitely when grown in culture. Some *cancer* cell populations can continue to divide indefinitely in culture and are therefore referred to as immortal. Senescence has been related to the continued activity of molecules that control *cell cycle* progression, and to the maintenance of the length of *telomeres* at the ends of *chromosomes*.

Stem cell A *cell* that can divide and produce two more stem cells, or divide and produce a stem cell and a differentiated cell, or divide and produce two differentiated cells. An embryonic stem may yield, after many cell divisions, all of the different specialized cells in the body. Such a stem cell is referred to as totipotent. In contrast, an adult stem cell, for instance, the hematopoietic stem cell in the bone marrow, may give rise to a multiple cell types in the blood. Such a stem cell with more restricted potential is referred to as a pluripotent or multipotent stem cell.

Synthetic biology The construction of organisms with new combinations of genetic regulatory elements. By analogy, these combinations of genetic regulatory elements are sometimes referred to as genetic circuits or genetic networks. Genetic circuits have been constructed that behave as switches, oscillators, timers, counters, clocks, logic processors, or sensors.

Systems biology The generation of data from complex, multiscale and dynamic genetic, biochemical, or metabolic pathways and networks of molecules using methods of experimental molecular biology, and the analysis of such data by techniques of mathematics, statistics, and computer science.

Telomeres The ends of *chromosomes*. The *DNA* at the ends of chromosomes contains repeated sequences (terminal restriction fragments, TRF) that are necessary for replicating DNA at the ends of chromosomes, and for maintaining the structural integrity of chromosomes.

Tissue architecture A spatially organized group of several types of differentiated cells. The different cell types may be organized into layers such as in the skin, or organized into structures such as branches of the lung, or organized into capped tubes such as the crypts of the colon.

Tumor See *Cancer*.

Tumor progression Also referred to as tumor evolution. A series of changes from normal *cells* to progeny cells that have a higher probability of dividing (hyperplasia) to cells that have an abnormal appearance (dysplastic) to cells that move out of their usual tissue location into adjacent tissue (invasive). A growth of cells that is not invasive is referred to as a benign neoplasm, whereas a growth of cells that is invasive is referred to as a malignant neoplasm. The term "cancer" usually refers to malignant neoplasm. Cells that form a tumor at a location distant from their original site are referred to as a metastatic tumor.

Virus An intracellular parasite of *cells*. There are viruses of bacteria and of higher cells, including mammalian cells. They replicate within cells and can be transferred between cells. The extracellular forms contain genetic material (*DNA* or *RNA*), *proteins*, and some contain membranes. Within cells, the viral genetic material may subvert the machinery of the host cells and alter the host cell's properties. Some RNA viruses, called retroviruses, can produce a DNA copy. The DNA copy can integrate into the host cell DNA and replicate along with the DNA of the host cell once per *cell cycle*. The integrated DNA copy of the virus may also produce RNA copies that get incorporated into new virus particles, which in turn can infect new cells.

Mathematical Glossary for Biologists

Cross-references to other glossary terms are *italicized*

Abel's equation One of the classical functional equations of the Calculus. For a supercritical *branching process*, the characteristic function of the limit *random variable W* equal to the standardized particle count satisfies Abel's Eq. (3.19).

Age-dependent branching process A *branching process* in which the lifetimes of particles are nonnegative *random variables*. In the special case when the lifetimes are exponentially distributed, the number of particles existing in the process, as a function of time, is a time-continuous *Markov chain*.

Asymptotic behavior Behavior of a time-dependent process (or a biological or physical phenomenon) after a sufficiently long time.

Backward approach Decomposition of the *branching process* into subprocesses started by direct progeny of the ancestor. By the branching property (a form of *self-recurrence*) these latter are distributed identically as the whole process. This decomposition provides means to derive recurrent relationships or equations for the distributions of the process.

Bellman–Harris branching process A *branching process* in which the lifetimes of particles are nonnegative *random variables* (age-dependent process) and the progeny is born exactly at the moment of the death of the parent.

Bienaymé–Galton–Watson branching process See *Galton–Watson branching process*.

Branching diffusion process A *branching process*, with a continuum *type space*, in which the type of the particle is defined as its position in a subset of real numbers (or points in higher-dimensional space) and the transitions in the type space are translations by a real-valued *random variable* (or a vector), with special rules on the boundary. The type may be understood as a spatial coordinate of the particle.

Branching process A random collection of individuals (particles, objects, cells), proliferating according to rules involving various degrees of randomness of the lifelength and the number of progeny of an individual. The unifying principle is the so-called branching property, which states that the lifelenght and type of progeny of a newborn particle, conditional on the current state of the process, are independent of any characteristics of other particles present at this time or in the future. The branching property is a form of *self-recurrence*.

Branching random walk A *branching process*, with a *denumerable* type space, in which the type of the particle is defined as its position in the set of integers (or nonnegative integers) and the transitions in the type space are translations by an integer *random variable*, with special rules on the boundary. An example is the process of *gene amplification* in proliferating *cells*. In this process the type of cell is the number of copies of a *gene* present in the cell's *DNA*. Progeny cells may gain or lose copies of this gene, inherited from the parent cell. So, if the number of gene copies in the parent is equal to i, then in the progeny it may be equal to $i-1$, i or $i+1$.

Càdlàg path Function of time continuous from the right and bounded from the left of each point (French: **c**ontinue **à d**roite, **l**imitée **à g**auche).

Chapman-Kolmogorov equation Fundamental relationship governing the time evolution of *Markov chains*. It is represented in various forms, e.g. $P(s+t) = P(s)P(t)$ or $P_{ij}(s+t) = \sum_k P_{ik}(s)P_{kj}(t)$, where $P(s) = (P_{ij}(s))$ is the matrix (finite or infinite) of transition probabilities between states, $P_{ij}(s) = \mathrm{P}[X_{t+s} = j | X_t = i]$. Intuitively, to calculate the probability of the chain moving from i to j in time $t+s$, it is necessary to add the probabilities of moving from i to k in time t, and from k to j in time s, over all states k.

Criticality *Branching process* is critical if the expected (mean) count of progeny of a particle is equal to 1. It is supercritical, if the mean count of progeny of a particle is greater than 1 and subcritical if it is less than 1. This classification leads to profound differences in *asymptotic properties* of the process. In particular, critical processes behave in a counterintuitive way since they become *extinct* with probability 1, while the expected number of particles stays constant.

Denumerable A set is called denumerable (or countable) if it is infinite but its elements can be indexed by nonnegative integers. Other categories of infinite sets include continuum, i.e., a set the elements of which can be indexed by real numbers from an interval. Set of all rational numbers (ratios of integers) is countable, set of all infinite sequences of zeros and ones is a continuum (since such sequences are just binary expansions of real numbers from the [0, 1] interval).

Exponential steady state For idealized populations growing without spatial or nutritional constraints, the condition in which the number of individuals increases or decreases exponentially, while the proportions of individuals in distinct age classes and any other identifiable categories remain constant. Usually attained *asymptotically*.

Extinction The event of all particles (individuals) of the *branching process* dying out.

Forward approach An approach dual to the backward approach, easiest to explain for the *Galton–Watson branching process*. Particles existing in generation t of the process are traced to their parents in generation $t-1$. Therefore, if the number Z_{t-1} of particles in generation $t-1$ is known, the number Z_t of particles in generation t is equal to the sum $Z_t = X_1 + X_2 + \ldots + X_{Z_{t-1}}$, where X_k is the number of progeny of the k-th out of Z_{t-1} particles of generation $t-1$. This leads to a recurrence for the *pgf*'s of the particle counts.

Galton–Watson branching process Arguably, the simplest *branching process*. It evolves in discrete time measured by nonnegative integers. At time 0, an ancestor individual (particle, cell, object) is born. At time 1, the ancestor dies, producing a random number of progeny. Each of these becomes an ancestor of an independent subprocess, distributed identically as the whole process. This definition implies that the numbers of progeny produced by each particle ever existing in the process are independent identically distributed *random variables* and that all particles live for one time unit. Discrete time moments coincide with generations of particles. The number of particles existing in the Galton–Watson branching process, as a function of time, constitutes a time-discrete *Markov chain*.

Gelation In a model of aggregation of chemical molecules, the idealized process of infinite aggregation, resulting in disappearance of finite aggregates of molecules. In Macken and Perelson's branching model of aggregation, gelation is represented by escape of the *branching process* to infinity (possible only in the supercritical case).

Genealogies Branching (tree-like) graphs, usually random with respect to structure and branch lengths, representing ancestry of a sample of individuals from a *branching process* or, more generally, from an abstract or real-life population of molecules, *genes*, *cells* or other objects. The process of reducing the number of distinct ancestors of the sample, followed in the reverse time, is called coalescence.

Genetic distance Distance between biological organisms, computed based on genetic characteristics. An example is the distance between relevant subsequences of *DNA* of the two individuals, computed as the number of nucleotides different in these two individuals (number of mismatches). For example, if in individual 1 the DNA sequence is *ATGGACGA* while in individual 2 it is *ATcGgCGt*, then the genetic distance is equal to 3.

iid Independent, identically distributed (*random variables*). The most frequently encountered assumption concerning a family of random variables. Makes proofs of theorems easier, when it can be assumed. In statistics, the so-called random samples are assumed to be iid.

Instability of branching processes The fact that, as time tends to infinity, the *branching process* either becomes *extinct* or infinitely large. Instability is due to the independence assumptions inherent in the definition of a branching process (i.e., that the number of progeny and lifelength of a newborn particle, conditional on the current state of the process are independent of any characteristics of other particles present at this time or in the future).

Jagers–Crump–Mode process The general *branching process*. The difference with respect to the classical branching processes, such as the *Galton–Watson branching process* or the *Bellman–Harris branching process* is that in the general process, the progeny may be produced before the death of the individual. The ages at which the individual begets progeny are random. Also, the type space may be of a very general form. The theory developed for general processes, allows finding distributions of the process counted by random characteristics, i.e. of the weighted counts of events associated with a desired subclass of individuals (e.g. the number of first-born progeny of all individuals born after January 1, 1980, etc.).

Kolmogorov theorem In the theory of *stochastic processes*, a fundamental result ensuring the existence of the stochastic process, given that for all finite collections of times, there exist joint distributions of *random variables* being the values of the process at these times. These finite-dimensional distributions have to satisfy consistency conditions.

Linear-fractional case An important case of the *Galton–Watson branching process*, in which the number of progeny of an individual is a random variable with modified geometric distribution, i.e., $\mathrm{P}[X = 0] = 1 - bp/(1 - p)$ and $\mathrm{P}[X = k] = bp^k$, for $k = 1, 2, \ldots$. The name is derived from the fact that the *pgf* of such random variable is a ratio of two linear functions. In the linear-fractional case,

the number of particles existing at any time has a modified geometric distribution, with parameters, which can be explicitly computed.

Malthusian parameter For a *branching process*, a parameter α such that the number $Z(t)$ of particles present in the process, normalized by dividing it by $\exp(\alpha t)$, converges to a limit *random variable*, as time tends to infinity. The Malthusian parameter always exists for the supercritical processes, and is positive in this case.

Markov branching process A type of time-continuous *branching process*. At time 0, an ancestor individual (particle, cell, object) is born. The ancestor lives for time τ, which is an exponentially distributed *random variable*, and then the ancestor dies, producing a random number of progeny. Each of these becomes an ancestor of an independent subprocess, distributed identically as the whole process. The number of particles existing in the Markov branching process, as a function of time, is a time-discrete Markov chain (hence the name). Interestingly, if the Markov branching process is observed at times equal to multiples of a constant interval Δt, the numbers of particles at these observation times are distributed identically as in a *Galton–Watson branching process*.

Markov process A *stochastic process* with a limited memory (the Markov property). Intuitively, given the state of the process at time t, the future of the process depends only on this state and not on its states at times before t (time can be discrete or continuous). Mathematically,

$$\mathrm{P}[X_{t+s} \in A | X_u = x_u,\ 0 \le s \le t] = \mathrm{P}[X_{t+s} \in A | X_t = x_t],$$

where A is a subset of the state space of the process (space of values assumed by the process). The probability $P(s; x \to A) = \mathrm{P}[X_{t+s} \in A | X_t = x]$ is the transition probability from state x to set of states A, in time s. If the states of the process form a finite or *denumerable* set, then the process is called a Markov chain. In this case, it is possible to define a matrix (finite or infinite) of transition probabilities between states $P(s) = (P_{ij}(s))$, where $P_{ij}(s) = \mathrm{P}[X_{t+s} = j | X_t = i]$. For discrete time, $P(s) = P(1)^s$, where $P(1)$ is the single-step transition probability matrix. For continuous time (under some additional assumptions if the number of states is infinite), $P(s) = \exp(Qs)$, where Q is called the transition intensity matrix.

Martingale In the discrete time case, a *stochastic process*, having the property that its expected value at time $t + 1$, conditional on its values at all times before $t + 1$, is equal to the process value at time t. Mathematically, $\mathrm{E}(X_{t+1} | X_1, X_2, \ldots, X_t) = X_t$. Martingales, under some additional conditions, converge to limits (which are *random variables*). For this reason, proving that a process is a martingale allows an insight into its *asymptotic behavior*. Continuous-time martingales behave in a similar way, but are technically more involved.

Maximum likelihood Statistical methodology of estimating parameters of models, based on observations. It consists of expressing the probability of observations as a function of parameters. This function is known as the likelihood function, $L(\theta) = f_X(x; \theta)$, where $f_X(\cdot)$ is the density of the distribution of *random variable* X, x is the vector of observations of random variable (known), and θ is the vector

of parameters of the distribution (unknown). The values of parameters, which maximize $L(\theta)$ are called the maximum likelihood estimates of the parameters and are denoted $\hat{\theta}$.

Moments Expected values of powers of a *random variable* X. Absolute moments of order k (or k-th absolute moments), are defined as $\mathrm{E}(X^k)$, central moments as $\mathrm{E}\{[X-\mathrm{E}(X)]^k\}$, and factorial moments as $\mathrm{E}[X(X-1)(X-2)\cdots(X-k)]$. First absolute moment, $\mathrm{E}(X)$, represents the central tendency of the random variable, second central moment, $\mathrm{Var}(X) = \mathrm{E}\{[X-\mathrm{E}(X)]^2\}$, represents the dispersion of the random variable around the expected value.

Multitype Galton–Watson process (positive regular) Generalization of the usual (single-type) *Galton-Watson branching process*. It evolves in discrete time measured by nonnegative integers. Each individual belongs to one of a finite number of types. At time 0, an ancestor individual (particle, cell, object), of some type, is born. Processes started by individuals of different types are generally different. At time 1, the ancestor dies, producing a random number of progeny of various types. The distribution of progeny counts depends on the type of parent. Each of the first-generation progeny becomes an ancestor of an independent subprocess, distributed identically as the whole process (modulo ancestor's type). In the multitype process, asymptotic behavior depends on the matrix of expected progeny count. Rows of this matrix correspond to the parents types, and columns to the progeny types. The largest positive eigenvalue of this matrix (the *Perron–Frobenius* eigenvalue), is the *Malthusian parameter* of the process, provided the process is supercritical (the Perron–Frobenius eigenvalue larger than 1) and positive regular. This latter means that parent of any given type will have among its (not necessarily direct) descendents individuals of all possible types, with nonzero probability.

Parsimony method in phylogenetics A method of inferring the *phylogenetic tree*. In this method, taxonomic units are represented by their *DNA* sequences (most commonly, from the *mitochondrial genome*). The method looks for the tree that requires the minimum number of changes between the extant and inferred ancestral sequences. The outcome may be equivocal, and also, since the number of possible tree structures is extremely large, the optimal tree is frequently not found.

Perron–Frobenius theory Collection of results concerning eigenvalues and eigenvectors of positive (or nonnegative) matrices and operators. Important assumptions include irreducibility (positive regularity), i.e., a strict positivity of iterates of the matrix or operator. A generic result states the existence of a strictly positive simple eigenvalue dominating all other eigenvalues and of a corresponding strictly positive eigenvector. The importance of these results is that they lead to characterizations of the *asymptotic behavior* of iterates of positive matrices or operators, in the terms of dominant eigenvalues and eigenvectors. Mathematically, $m_0 M^i \sim \lambda^i \nu$, as $i \to \infty$, where M^i is the i-th iterate of the positive matrix M, m_0 is the initial vector of states, λ is the dominant positive eigenvalue and ν is the corresponding eignvector. Results of this type are important in mathematical population dynamics, including the theory of *branching processes*.

pgf *Probability generating function*.

Phylogenetic tree The set of ancestry relationships between extant (contemporary) taxonomic or demographic units (species, populations, haplotypes and other), usually in the form of a binary tree graph (at most three branches out of each node). The nodes of the phylogenetic tree represent extant and ancestral units, while the branches represent the intervals of evolutionary time separating them. Depending on the method of reconstruction, the graph may be rooted, i.e. having a uniquely defined common ancestor (and consequently, the direction of time specified in all branches), or unrooted (it is then sometimes called a network). The most commonly used methods of reconstruction are *parsimony*, distance matrix and *maximum likelihood*.

Poisson process One of the most important *stochastic processes*. Random collection of time points (epochs) having the properties of complete randomness (the counts of events in any two disjoint time intervals are independent), and stationarity (the probability of an event occurring in a short time interval $(t, t + dt)$ is equal to $\lambda dt + o(dt)$, where $o(dt)$ is small with respect to dt, i.e., $o(dt)/dt \to 0$ as $dt \to 0$). The constant λ is called the intensity of the process. The number N of epochs of the Poisson process in an interval of length t has Poisson distribution with parameter λt (i.e., $\mathrm{P}[N = n] = \exp(-\lambda t)(\lambda t)^n/n!$, for $n = 0, 1, 2, \ldots$), and the time intervals T between any two epochs have exponential distribution with parameter λ (i.e., the density of distribution of T is equal to $f_T(t) = \lambda \exp(-\lambda t)$, for $t \geq 0$).

Population genetic models Models of inheritance, *mutation* and selection of genetic material in populations of individuals. Classically, these models assume a constant number of individuals related to each other through common ancestry (Fisher-Wright model). Although very different from the *branching processes*, some of these models can be approximated by branching processes, e.g. when an expanding subpopulation of mutants arises within the large population. Such situation arises when some of genetic diseases are studied.

Positivity In general, the property of being positive. A matrix is positive if all elements of the matrix are positive; it is positive regular, if all elements are nonnegative and some power of the matrix is positive. If the matrix is a transition probability matrix of a Markov process, positive regularity means that there exist paths between all pairs of states of the process. Similarly, if the matrix is the mean progeny matrix of a multitype *branching process*, than positive regularity means that any particle has, among its descendants, particles of all types.

Probability generating function (*pgf*) Function $f_X(s)$ of a symbolic argument s, which is an equivalent of the distribution of a nonnegative-integer valued *random variable* X. If numbers $p_0, p_1, p_2, \ldots$ constitute the distribution of random variable X, i.e., $\mathrm{P}[X = k] = p_k$, then the pgf of random variable X is defined as $f_X(s) = \mathrm{E}(s^X) = \sum_{i=0}^{\infty} p_i s^i$, for $s \in [0, 1]$. Use of the pgf simplifies mathematical derivations involving nonnegative-integer random variables.

Quasistationarity State i_a of a *Markov chain* $X(t)$ is called absorbing, if the process cannot exit i_a, once i_a has been visited, i.e., $\mathrm{P}[X(t + s) \neq i_a | X(t) = i_a] = 0$. Under certain additional conditions, the probability of eventual absorption in state

i_a is equal to 1, i.e., $\mathrm{P}[\lim_{t\to\infty} X(t) = i_a] = 1$. Then the only *stationary* distribution of is the one that assigns probability 1 to state i_a. Since such distribution is not informative, it is usual to consider a distribution, which is stationary conditional on non-absorption. Such distribution, if it exists, is called the quasistationary distribution. Mathematically, $\tilde{\pi} = (\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \dots)$ is the quasistationary distribution, if $\mathrm{P}[X(t+s) = j | X(t+s) \neq i_a] = \tilde{\pi}_j$ (all j) provided $\mathrm{P}[X(t) = j | X(t) \neq i_a] = \tilde{\pi}_j$ (all j). An example of a quasistationary distribution is the limit distribution of the subcritical *branching process* conditional on non-extinction.

Random variable (*rv*) Intuitively, a numerical result of observation, which displays random variation. Mathematically, a random variable $X(\omega)$ is a function mapping the elements ω of a probability space Ω (space of outcomes of a random experiment) into the set of real numbers. For technical reasons, this function has to be measurable, i.e., the counter image of an interval through X has to be a measurable set of elements of Ω.

Random walk A time-discrete *Markov chain* $X(t)$, such that $X(t+1) = X(t) + U(t)$, where the integer *random variables* $U(t)$ are independent and identically distributed.

Recurrent state See *transient state*.

Renewal theory A branch of probability concerned with renewal processes. The renewal process is a collection of random time points (called renewals) such that the intervals between these points are independent, identically distributed *random variables*. A special case in which the intervals between renewals are exponentially distributed is the *Poisson process*.

rv *Random variable*

Self-recurrence Consider a random (*stochastic) process* $X(t)$, evolving from an initial value $X(0) = x_0$ on time interval $[0, \infty)$. Suppose that at some time t_0, the process is stopped and then restarted. Then, suppose that given the value $X(t_0) = x_0$, the continuation process on the interval $[t_0, \infty)$, which is a subprocess of the original process, is identical (it has the same distributions) as the original process shifted by t_0. A process with such property is called self-recurrent. Self-recurrence may be considered a rephrasing of a causality principle. It leads to recurrent relationships for a wide class of processes, including *Markov processes, renewal processes* and *branching processes*.

Stathmokinesis An experimental technique in which *cell* divisions are blocked, ideally without damage to cells. Cells traversing successive phases of their lives are accumulating in the pre-division state (mitosis). Time pattern of accumulation depends on the demography of the cell population and kinetic parameters of the cell cycle. Therefore, it is possible to estimate some of these parameters based on observed accumulation patterns.

Stationarity *Markov chain* $X(t)$ is said to be stationary, if its distribution over the state space is invariant in time (this distribution is called the stationary distribution). Mathematically, $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ is the stationary distribution, if $\mathrm{P}[X(t+s) = j] = \pi_j$ (all j) provided $\mathrm{P}[X(t) = j] = \pi_j$ (all j).

Stochastic process Intuitively, a function of time with a random component. Mathematically, a family of *random variables* parameterized by time. It has to satisfy

so-called measurability conditions, which prevent certain mathematical problems from occurring.

Transient state States of a *Markov chain* can be classified into transient and *recurrent*. For a recurrent state the probability of eventually returning to this state is equal to 1, while for a transient state there is a nonzero probability of never returning.

Type space A collection of possible particle types existing in a *branching process*. If there is more than one but finitely many types, the process is called multitype. If the type space is *denumerable* or continuous, the behavior of the branching process can differ considerably from the multitype case. An example is a *branching random walk*, in which the asymptotic behavior can be, for example, exponential multiplied by fractional power function, which does not occur in the finite case.

wp *With probability* (common abbreviation)

Yaglom's theorem Result stating that for subcritical *branching processes*, there exists a *quasistationary* distribution, conditional on non-*extinction*.

Yule process *Markov age-dependent branching process* in which a particle can have at most two progeny (the binary fission process). An important class of processes, since the *pgf* of the distribution of particle count can be explicitly found. Also, the Yule process frequently serves as a model for populations of proliferating *cells*, although by its definition it is limited to exponentially distributed cell lifetimes.

References

Abkowitz JL, Catlin SN, Guttorp P (1996) Evidence that hematopoiesis may be a stochastic process in vivo. Nat Med 2:190–197

Adami C, Chu J (2002) Critical and near-critical branching processes. Phy Rev E 66(1 Pt 1):011907

Afanasyev VI (1980) Limit theorems for a conditional random walk and some applications. Dissertation for the Candidate of Sciences degree, Moscow State University (in Russian)

Aldous D (1993) Tree-based models for random distribution of mass. J Stat Phy 73:625–641

Aldous DJ, Krikun MA, Popovic L (2011) Five statistical questions about the tree of life. Syst Biol 60:318–328

Alexandersson M (1999) Branching processes and cell populations. Unpublished Ph.D. thesis. Department of Mathematical Statistics, Chalmers University, Göteborg, Sweden

Alsmeyer G, Gutierrez C, Martínez R (2011) Limiting genotype frequencies of Y-linked genes through bisexual branching processes with blind choice. J Theor Biol 275:42–51

Alt W, Tyson JJ (1987) A stochastic model of cell division (with application to fission yeast). Math Biosci 83:1–29

Angerer WP (2001) An explicit representation of the Luria–Delbrück distribution. J Math Biol 42:145–174

Aout M (2006) A branching cell population model allowing for cell death. J Math Stat 2:373–377

Arino O, Kimmel M, Webb GF (1995) Mathematical modeling of the loss of telomere sequences. J Theor Biol 177:45–57

Arino O, Kimmel M (1987) Asymptotic analysis of a cell cycle model based on unequal division. SIAM J Appl Math 47:128–145

Arino O, Kimmel M (1991) Asymptotic behavior of nonlinear semigroup describing a model of selective cell growth regulation. J Math Biol 29:289–314

Arino O, Kimmel M, Zerner M (1991) Analysis of a cell population model with unequal division and random transition. In: Arino O, Axelrod DE, Kimmel M (eds) Mathematical population dynamics. Marcel Dekker, New York, pp 3–12

Arino O, Kimmel M (1993) Comparison of approaches to modeling of cell population dynamics. SIAM J Appl Math 53:1480–1504

Arking R (1998) Biology of aging. Sinauer, Sunderland

Arkus N (2005) A mathematical model of cellular apoptosis and senescence through the dynamics of telomere loss. J Theor Biol 235:13–32

Asmussen S, Hering H (1983) Branching processes. Birkhäuser, Boston

Asteris G, Sarkar S (1996) Bayesian procedures for the estimation of mutation rates from fluctuation experiments. Genetics 142:313–326

Athreya KB, Ney PE (2004) Branching processes. Dover Publications, Mieola

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

Athreya KB (2012a) Coalescence in recent past in rapidly growing populations. Stoch Proc Appl 122:3757–3766
Athreya KB (2012b) Coalescence in critical and subcritical Galton–Watson branching processes. J Appl Probab 49:627–638
Axelrod DE, Baggerly KA, Kimmel M (1994) Gene amplification by unequal sister chromatid exchange: probabilistic modeling and analysis of drug resistance data. J Theor Biol 168:151–159
Axelrod DE, Gusev Y, Gamel JW (1997) Ras-oncogene transformed and non-transformed cell population are each heterogeneous but respond differently to the chemotherapeutic drug cytosine arabinoside (Ara-C). Cancer Chemother Pharmacol 39:445–451
Axelrod DE, Gusev Y, Kuczek T (1993) Persistence of cell cycle times over many generations as determined by heritability of colony sizes of *ras* oncogene-transformed and non-transformed cells. Cell Prolif 26:235–249
Axelrod DE, Haider FR, Tate AC (1986) Distribution of interdivisional times in proliferating and differentiating Friend murine erythroleukaemia cells. Cell Tissue Kinet 19:547–556
Axelrod DE, Kuczek T (1989) Clonal heterogeneity in populations of normal cells and tumor cells. Comput Math Appl 18:871–881
Azevedo RB, Leroi AM (2001) A power law for cells. Proc Natl Acad Sci U S A 98:5699–5704
Baake E, Georgii HO (2007) Mutation, selection, and ancestry in branching models: a variational approach. J Math Biol 54:257–303
Baggerly KA, Kimmel M (1995) Emergence of stable DNA repeats from random sequences under unequal sister chromatid exchange. In: Lakshmikantham V (ed) Proceedings of the 1st World Congress of Nonlinear Analysts, Tampa, Florida, August 1992. Walter de Gruyter, Berlin, pp 3409–3418
Baird SJ, Barton NH, Etheridge AM (2003) The distribution of surviving blocks of an ancestral genome. Theor Popul Biol 64:451–471
Ball F, Neal P (2010) Applications of branching processes to the final size of SIR epidemics. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin. pp 207–223
Bansaye V (2008) Proliferating parasites in dividing cells: Kimmel's branching model revisited. Ann Appl Probab 18:967–996
Bartlett MS (1984) Recent investigations involving stochastic population models. Adv Appl Probab 16:449–470
Barton NH, Briggs DEG, Eisen JA, Goldstein DB (2007) Evolution. Cold Spring Harbor Laboratory Press. Cold Spring Harbor
Bat O, Kimmel M, Axelrod DE (1997) Computer simulation of expansions of DNA triplet repeats in the fragile X syndrome and Huntington's disease. J Theor Biol 188:53–67
Beaudry AA, Joyce GF (1992) Directed evolution of an RNA enzyme. Science 257:635–641
Beil M, Luck S, Fleischer F, Portet S, Arendt W, Schmidt V (2009) Simulating the formation of keratin filament networks by a piecewise-deterministic Markov process. J Theor Biol 256: 518–532
Berndtsson B, Jagers P (1979) Exponential growth of a branching process usually implies stable age distribution. J Appl Probab 16:651–656
Bertolusso R, Kimmel M (2012) Modeling spatial effects in early carcinogenesis: stochastic versus deterministic reaction-diffusion systems. Math Model Nat Phenom 7:245–260
Bertuzzi A, Gandolfi A (1983) Recent views on the cell cycle structure. Bull Math Biol 45:605–616
Bertuzzi A, Gandolfi A, Giovenco MA (1981) Mathematical models of the cell cycle with a view to tumor studies. Math Biosci 53:159–188
Bichsel M, Barbour AD, Wagner A (2010) The early phase of a bacterial insertion sequence infection. Theor Popul Biol 78:278–288
Biggins JD (1977) Chernoff's theorem in the branching random walk. J Appl Probab 14:630–636
Biggins JD (1995) The growth and spread of the general branching random walk. Ann Appl Probab 5:1008–1024

Biggins JD (1997) How fast does a general branching random walk spread? In: Jagers P, Athreya K (eds) Classical and modern branching processes. The IMA volumes in mathematics and its applications, 84. Springer, Berlin, pp 19–39

Biggins JD, Kyprianou AE (1996) Branching random walk: Seneta-Heyde norming. In: Chauvin B, Cohen S, Rouault A (eds) Trees. Progress in Probability, 40. Birkhäuser, Basel, pp 31–49

Biggins JD, Lubachevsky BD, Shwartz A, Weiss A (1991) A branching random walk with a barrier. Ann Appl Probab 1:573–581

Birky CW, Skavaril RV (1984) Random patitioning of cytoplasmic organelles at cell division: the effect of organelle and cell volume. J Thor Biol 106:441–447

Blackburn EH (1991) Structure and function of telomeres. Nature 350:569–573

Block GL, Allen LJ (2000) Population extinction and quasi-stationary behavior in stochastic density-dependent structured models. Bull Math Biol 62:199–228

Bobrowski A, Kimmel M (1999) Asymptotic behaviour of an operator exponential related to branching random walk models of DNA repeats. J Biol Syst 7:33–43

Bobrowski A, Kimmel M (2004) Asymptotic behavior of joint distributions of characteristics of a pair of randomly chosen individuals in discrete-time Fisher-Wright models with mutations and drift. Theor Popul Biol 66:355–367

Borovkov KA, Vatutin VA (1977) Reduced critical branching processes in random environment. Stoch Proc Appl 77:225–240

Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA (2010) Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 107:18445–18550. Commentatry: Regoes, R. Proc Natl Acad Sci U S A 107:1824–1842. Population genetics meets cancer genomics

Breiman L (1968) Probability. Addison-Wesley, Reading

Briggs AW, Good JM, Green RE, Krause J et al (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325:318–321

Brooks RF, Bennett DC, Smith JA (1980) Mammalian cell cycles need two random transitions. Cell 19:493–504

Brown PC, Beverly SM, Schimke RT (1981) Relationship of amplified dihydrofolate reductase genes to double minute chromosomes in unstably resistant mouse fibroblasts cell lines. Mol Cell Biol 1:1077–1083

Bühler WJ (1992) A multitype branching model for populations with formation of subunits. In: Schach S et al. (eds) Data analysis and statistical inference: Festschrift in honour of Prof. Dr. Friedhelm Eicker. Bergisch Gladbach, Verlag Josef Eul, pp 185–190

Campbell RB (2003) A logistic branching process for population genetics. J Theor Biol 225:195–203

Campos PR (2004) Fixation of beneficial mutations in the presence of epistatic interactions. Bull Math Biol 66:473–486

Caskey CT, Pizutti A, Fu Y-H, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–789

Chen R, Hyrien O, Noble M, Mayer-Proschel M (2011) A composite likelihood approach to the analysis of longitudinal clonal data on multitype cellular systems under an age-dependent branching process. Biostatistics 12:173–191

Chinnery PF, Turnbull DM (1999) Mitochondrial DNA and disease. Lancet 354:S17–S21

Ciampi A, Kates L, Buick R, Kriukov Y, Till JE (1986) Multi-type Galton-Watson process as a model for proliferating human tumour cell populations derived from stem cells: Estimation of stem cell self-renewal probabilities in human ovarian carcinomas. Cell Tissue Kinet 19:129–140

Cohn H, Klebaner F (1986) Geometric rate of growth in Markov Chains with applications to population-size-dependent models with dependent offspring. Stoch Anal Appl 4:283–308

Coldman AJ (1987) Modeling resistance to cancer chemotherapeutic agents. In: Thompson JR, Brown BW (eds) Cancer modeling. Marcel Dekker, New Year, pp 315–364

Coldman AJ, Goldie AJ (1983) A model for the resistance of tumor cells to cancer chemotherapeutic agents. Math Biosci 65:291–307

Coldman AJ, Goldie JH (1985) Role of mathematical modeling in protocol formulation in cancer chemotherapy. Cancer Treat Rep 69:1041–1048
Coldman AJ, Goldie JH (1986) A stochastic model for the origin and treatment of tumors containing drug resistant cells. Bull Math Biol 48:279–292
Coldman AJ, Goldie JH, Ng V (1985) The effect of cellular differentiation on the development of permanent drug resistance. Math Biosci 74:177–198
Conway JM, Coombs D (2011) A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients. PLoS Comput Biol 7(4):e1002033
Cooper S (1979) A unifying model for the G1 period in prokaryotes and eukaryotes. Nature 280: 17–19
Cooper S (1984) The continuum model as a unified description of the division cycle of eukaryotes and prokaryotes. In: Nurse P, Streiblova E (eds) The microbial cell cycle. CRC, Boca Raton, pp 8–27
Cooper S (1991) Bacterial growth and division: biochemistry and regulation of prokaryotic and eukaryotic division cycles. Academic Press, San Diego
Cottrell D, Swain PS, Tupper PF (2012) Stochastic branching-diffusion models for gene expression. Proc Natl Acad U S A 109:9699–9604
Counter CM, Avilion AA, Lefeuvre CE, Stewart and others (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J 11:1921–1929
Cowan R (1985) Branching process results in terms of moments of the generation-time distribution. Biometrics 41:681–689
Cowan R, Culpin D (1981) A method for the measurement of variability in cell lifetimes. Math Biosci 54:249–263
Cowan R, Morris VB (1986a) Cell population dynamics during the differentiation phase of tissue development. J Theor Biol 122:205–224
Cowan R, Staudte R (1986b) The bifurcating autoregression model in cell lineage studies. Biometrics 42:769–783
Crump KS (1970a) On systems of renewal equations. J Math Anal Appl 30:425–434
Crump KS (1970b) Migratory populations in branching processes. J Appl Probab 7:565–572
Crump KS, Mode CJ (1969) An age-dependent branching process with correlations among sister cells. J Appl Probab 6:205–210
Crump KS, Hoel DG (1974) Mathematical models for estimating mutation rates in cell populations. Biometrika 61:237–252
Cyran KA, Kimmel M (2010) Alternatives to the Wright–Fisher model: the robustness of mitochondrial Eve dating. Theor Popul Biol 78:165–172
Czerniak B, Herz F, Wersto RP, Koss LG (1992) Asymmetric distribution of oncogene products at mitosis. Proc Natl Acad Sci U S A 89:4860–4863
Danesh K, Durrett R, Havrilesky L, Myers E (2012) A branching process model of ovarian cancer. J Theor Biol 314:10–15
Darzynkiewicz Z, Carter S, Kimmel M (1984) Effects of [3H]Udr on the cell-cycle progression of L1210 cells. Cell Tissue Kinet 17:641–655
Darzynkiewicz Z, Crissman H, Traganos F, Steinkamp J (1982) Cell heterogeneity during the cell cycle. J Cell Physiol 113:465–474
Darzynkiewicz Z, Evenson DP, Staiano-Coico L, Sharpless TK, Melamed ML (1979) Correlation between cell cycle duration and RNA content. J Cell Physiol 100:425–438
Darzynkiewicz Z, Traganos F, Kimmel M (1986) Assay of cell cycle kinetics by multivariate flow cytometry using the principle of stathmokinesis. In: Gray JE, Darzynkiewicz Z (eds) Techniques in cell cycle analysis. Humana, Clifton, pp 291–336
Darzynkiewicz Z, Traganos F, Kimmel M, Myc A (1990) Kinetic and metabolic heterogeneity of cells in the cell cycle. In: Tautu P (ed) Stochastic modeling in biology. Relevant mathematical concepts and recent applications. World Scientific Publishing, Singapore

Darzynkiewicz Z, Traganos F, Xue S, Stainao-Coico L, Melamed MR (1981) Rapid analysis of drug effects on the cell cycle. Cytometry 1:279–286
Dawson D, Perkins E (1991) Historical processes. Mem Am Math Soc 93(454)
Dawson DA, Hochberg KJ (1991) A multilevel branching model. Adv Appl Probab 23:701–715
Day RS (1986a) Treatment sequencing, asymmetry, and uncertainty: protocol strategies for combination chemotherapy. Cancer Res 46:3876–3885
Day R (1986b) A branching process model for heterogeneous cell populations. Math Biosci 78: 73–90
Demetrius L, Schuster P, Sigmund K (1985) Polynuclotide evolution and branching processes. Bull Math Biol 47:239–262
Demos JP (1982) Entertaining Satan: witchcraft and the culture of early New England. Oxford University, New York
DeVita VT, Lawrence TS, Rosenberg SA (eds) (2010) Cancer: principles and practice of oncology. Walters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia
Dibrov BF, Zhabotinsky AM, Neyfakh YA, Orlova MP, Churikova LI (1985) Mathematical model of cancer chemotherapy: periodic schedules of phase specific cytotoxic agent administration increasing the selectivity of therapy. Math Biosci 73:1–31
Dibrov BF, Zhabotinsky AM, Neyfakh YA, Orlova MP, Churikova LI (1983) Optimal scheduling for cell synchronization by cycle-specific blockers. Math Biosci 66:167–185
Doetsch G (1974) Introduction to the theory and application of the Laplace transformation. Springer-Verlag, New York
Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge University, Cambridge
Durrett R (1978) The genealogy of critical branching processes. Stoch Process Appl 8:101–116
Durrett R, Foo J, Leder K, Mayberry J, Michor F (2010) Evolutionary dynamics of tumor progression with random fitness values. Theor Popul Biol 78:54–66
Durrett R, Mayberry J (2011) Traveling waves of selective sweeps. Ann Appl Prob 21:699–744
Durrett R, Schweinsberg J (2005) Power laws for family sizes in a duplication model. Ann Appl Probab 33:2094–2126
Etheridge AM (1993) Conditioned superprocesses and a semilinear heat equation. In: Cinlar E, Chung KL, Sharpe MJ (eds) Seminar on stochastic processes: progress in probability, 33. Birkhäuser, Boston, pp 89–99 (Seattle, WA, 1992)
Falahati A (1999) Two-sex branching populations. Doctoral thesis. Department of Mathematical Statistics, Chalmers University. Doctoral Dissertations Series no. 1493. Göteborg, Sweden
Fathi-Vojargah B, Moradi M (2011) Period dependent branching process and its applications in epidemiology. Infect Genet Evol 11:1225–1228
Fearn DH (1972) Galton-Watson processes with generation dependence. Proceedings of the Sixth Berkley Symposium on Mathematical and Statistical Probabilitity 4, pp 159–172
Fearn DH (1976) Supercritical age dependent branching process with generation dependence. Ann Probab 4:27–37
Feller W (1968) An introduction to probability and its applications, vol 1, 3rd edn. Wiley, New York
Feller W (1971) An introduction to probability and its applications, vol 2, 2nd edn. Wiley, New York
Fleischmann K, Siegmund-Schultze R (1977) The structure of the reduced critical Galton–Watson processes. Mathematische Nachrichten 79:233–241
Fleischmann K, Vatutin VA (1999) Reduced subcritical Galton–Watson processes in random environment. Adv Appl Probab 31:1–24
Foster PL (2006) Methods for determining spontaneous mutation rates. Methods Enzymol 409:195–213
Fujimagari T (2007) On a periodic-like behavior of a delayed density-dependent branching process. Math Biosci 206:128–133
Gawel B, Kimmel M (1996) Iterated Galton–Watson process. J Appl Probab 33:949–959

Geli P, Andersson M, Svensson A, Andersson DI (2009) A multi-type branching model with varying environment for bacterial dynamics with postantibiotic effect. J Theor Biol 256:58–64
Gernhard T (2008) New analytic results for speciation times in neutral models. Bull Math Biol 70:1082–1097
Ghosh A, Haccou P (2010) Quantifying stochastic introgression processes with hazard rates. Theor Popul Biol 77:171–180
Gillespie JH (1986) Variability of evolutionary rates of DNA. Genetics 113:1077–1091
Glauche I, Moore K, Thielecke L, Horn K, Loeffler M, Roeder I (2009) Stem cell proliferation and quiescence—two sides of the same coin. PLoS Comput Biol 5:e1000447
Goldie AJ (1982) Drug resistance and chemotherapeutic strategy. In: Owens AH, Coffey DS, Baylin SB (eds) Tumor cell heterogeneity. Academic Press, New York, pp 115–125
Goldie JH, Coldman AJ (1979) A mathematical model for relating the drug sensitivity of tumors to their spontaneous mutation rate. Cancer Treat Rep 63:1727–1733
Goldie JH, Coldman AJ (1984) The genetic origin of drug resistance in neoplasms: implications for systemic therapy. Cancer Res 44:3643–3653
Goldie JH, Coldman AJ, Gudauskas GA (1982) Rationale for the use of alternating non-cross-resistant chemotherapy. Cancer Treat Rep 66:439–449
González M, Guite'rrex C, Martínez R (2010) Parametric inference for Y-linked gene branching models: expectation-maximization method. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 191–204
González M, Hull DM, Martínez R, Mota M (2006) Bisexual branching processes in a genetic context: the extinction problem for Y-linked genes. Math Biosci 202:227–247
González M, Martínez R, Mota M (2008) Bisexual branching processes in a genetic context: rates of growth for Y-linked genes. Math Biosci 215:167–176
González M, Martínez R, Slavtchova-Bojkova M (2010) Time to extinction of infrectious diseases throught age-dependent branching process. In: González M et al. (eds) Workshop on Branching Processes and Their Applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 241–256
González M, Molina M (1996) On the limit behaviour of a superadditive bisexual Galton–Watson branching process. J Appl Probab 33:960–967
Green J (1981) Modelling yeast cell-growth using stochastic branching processes. Appl Probab 18:799–808
Green RE, Malaspinas AS, Krause J, Briggs, AW et al (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134:416–426
Greider CW (1996) Telomere length regulation. Annu Rev Biochem 65:337–365
Greider CW, Blackburn EH (1996) Telomeres, telomerase and cancer. Sci Am 274(2):92–97
Griffiths R, Pakes A (1988) An infinite-alleles version of the simple branching process. Adv Appl Probab 20:489–524
Griffiths RC, Tavaré S (1999) The ages of mutations in gene trees. Ann Appl Probab 9:567–590
Guo X, Bernard A, Orlandob DA, Haase SB, Harteminka AJ (2013) Branching process deconvolution algorithm reveals a detailed cell-cycle transcription program. Proc Natl Acad Sci U S A 110(10):E968–E977
Gupta PB, Fillmore AG, Jiang G, Shapire, SD et al (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146:633–644. (Comment: McCarthy N, Nature reviews. Cancer 11:690)
Gusev Y, Axelrod DE (1995) Evaluation of models of inheritance of cell cycle times: computer simulation and recloning experiments. In: Arino O, Axelrod DE, Kimmel M (eds) Mathematical population dynamics: analysis of heterogeneity. vol 2. Carcinogenesis and cell. pp 97–116
Guttorp P (1991) Statistical inference for branching processes (Wiley Series in probability and mathematical statistics). Wiley, New York
Gyllenberg M (1986) The size and scar distributions of the yeast *Saccharomyces cerevisiae*. J Math Biol 24:81–101

Haccou P, Jagers P, Vatutin V (2005) Branching processes. Cambridge University Press, Cambridge
Haccou P, Vatutin V (2003) Establishment success and extinction risk in autocorrelated environments. Theor Popul Biol 64:303–314
Haldeman C, Beggs JM (2005) Critical branching captures activity in living neural networks and maximizes the number of metastable states. Phys Rev Lett 94:058101
Harley CB (1991) Telomere loss: mitotic clock or genetic time bomb? Mutat Res 256:271–282
Harley CB, Goldstein S (1980) Retesting the commitment theory of cellular aging. Science 207:191–193
Harnevo LE, Agur Z (1991) The dynamics of gene amplification described as a multitype compartmental model and as a branching process. Math Biosci 103:115–138
Harnevo LE, Agur Z (1992) Drug resistance as a dynamic process in a model for multistep gene amplification under various levels of selection stringency. Cancer Chemother Pharmacol 30:469–476
Harnevo LE, Agur Z (1993) Use of mathematical models for understanding the dynamics of gene amplification. Mutat Res 292:17–24
Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Nat Acad Sci U S A 95:1961–1967
Harris TE (1963) The theory of branching processes. Springer-Verlag, Berlin
Hartfield M, Otto SP (2011) Recombination and hitchhiking of deleterious alleles. Evolution 65:2421–2434
Hasegawa M, Horai S (1990) Time of the deepest root for polymorphism in human mitochondrial DNA. J Mol Evol 32:37–42
Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204–211
Hastings PJ, Bull HJ, Klump JR, Rosenberg SM (2000) Adaptive amplification: an inducible chromosomal instability mechanism. Cell 103:723–731
Hautphenne S (2009) An algorithm look at phase-controlled branching processes. Thesis, Université Libre de Bruxelles
Hautphenne S, Latoouche G, Remiche M-A (2011) Algorithmic approach to the extinction probability of branching processes. Methodol Comput Appl Probab 13:171–192
Heinrich V, Stange J, Dickhaus T, Imkeller P, Krüger U, Bauer S, Mundios S, Robinson PN, Hecht J, Krawitz PM (2011) The allele distribution in next-generation sequencing data sets is accurately described as the result of a stochastic branching process. Nucleic Acids Res 40:2426–2431
Heizmann D (2010) Time to extinction in a two-host interaction model for the macroparasite *Echinococcus granulosus*. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 257–266
Hejblum G, Costagiola D, Valleron A-J, Mary J-Y (1988) Cell cycle models and mother-daughter correlation. J Theor Biol 131:255–262
Hoel DG, Crump KS (1974) Estimating the generation-time distribution of an age-dependent branching process. Biometrics 30:125–135
Holmes EC (2009) The evolution and emergence of RNA viruses. Oxford University, Oxford
Horn M, Loeffler M, Roeder I (2008) Mathematical modeling of genesis and treatment of chronic myeloid leukemia. Cells Tissues Organs 188:236–247
Hyrien O, Chen R, Zand MS (2010) An age-dependent branching process model for the analysis of CFSE-labeling experiments. Biol Direct 5:41
Hyrien O, Dietrich J, Noble M (2010) Mathematical and experimental approaches to identify and predict the effects of chemotherapy on neuroglial precursors. Cancer Res 70:10051–10059
International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
Iwasa Y, Michor F, Nowak MA (2003) Evolutionary dynamics of escape from biomedical intervention. Proc Royal Soc London B 270:2573–2578

Iwasa Y, Michor F, Nowak MA (2004) Evolutionary dynamics of invasion and escape. J Theor Biol 226:205–214
Jacob C (2010) Branching processes: their role in epidemiology. Int J Environ Res Public Health 7:1186–1204
Jacob C, Maillard-Teyssier L, Denis JB, Bidot C (2010) A branching process approach for the propagation of the Bovine-Spongiform Encephalopathy in Great Britain. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 225–240
Jagers P (1975) Branching processes with biological applications. Wiley, London
Jagers P (1983) Stochastic models for cell kinetics. Bull Math Biol 45:507–519
Jagers P (1991) The growth and stabilization of populations. Stat Sci 6:269–283
Jagers P (1992) Stabilities and instabilities in population dynamics. J Appl Probab 29:770–780
Jagers P (1997) Coupling and population dependence in branching processes. Ann Appl Probab 7:281–298
Jagers P (2001) The deterministic evolution of general branching populations. In: State of the art in probability and statistics (Leiden, 1999), pp 384–398. IMS Lecture notes monograph series 36, Institute of Mathematical Statistics, Beachwood, OH
Jagers P, Klebaner F (2003) Random variation and concentration effects in PCR. J Theor Biol 224:299–304
Jagers P, Klebaner F, Sagitov S (2007) On the path to extinction. Proc Natl Acad Sci U S A 104:6107–6111
Jagers P, Nerman O (1996) The asymptotic composition of supercritical multi-type branching populations. In: Mar Y (ed) Séminaire de Probabilités. Lecture notes in mathematics. Springer-Verlag, Berlin
Jagers P, Norrby K (1974) Estimation of the mean and variance of cycle times in cinemicrographically recorded cell populations during balanced exponential growth. Cell Tissue Kinet 7:201–211
Jagers P, Sagitov S (2008) General branching processes in discrete time as random trees. Bernoulli 14:949–962
Joffe A, Letac G (2006) Multitype linear fractional branching processes. J Appl Probab 43:1091–1106
Joffe A, Waugh W (1982) Exact distributions of kin numbers in a Galton–Watson process. J Appl Probab 19:767–775
Joffe A, Waugh W (1986) Exact distributions of kin numbers in a Galton–Watson process. In: Janssen J (ed) Semi-Markov models. Plenum Press, New York, pp 397–405
Joffe A, Waugh W (1986) The kin number problem in a multitype Galton–Watson population. J Appl Probab 22:37–47
Jones G (2011a) Calculations for multi-type age-dependent binary branching processes. J Math Biol 63:33–56
Jones G (2011b) Tree models for macroevolution and phylogenetic analysis. Syst Biol 60:735–746
Jones ME (1994) Luria–Delbrück fluctuation experiments: accounting simultaneously for plating efficiency and differential growth rate. J Theor Biol 166:355–363
Jones ME, Thomas SM, Rogers A (1994) Luria–Delbrück fluctuation experiments: design and analysis. Genetics 136:1209–1216
Jones ME, Wheldrake J, Rogers A (1993) Luria-Delbrück fluctuation analysis: estimating the Poisson parameter in a compound Poisson distribution. Comput Biol Med 23:525–534
Jones RB, Lumpkin CK Jr, Smith JR (1980) A stochastic model for cellular senescence. Part I. Thoeretical considerations. J Theor Biol 86:581–592
Joyce GF (1992) Directed molecular evolution. Sci Am 267(Dec):90–97
Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56:18–32
Kaufman RJ, Brown PC, Schimke RT (1981) Loss and stabilization of amplified dihydrofolate reductase genes in mouse sarcoma S-180 cell lines. Mol Cell Biol 1:1084–1093

Kendal WS, Frost P (1988) Pitfalls and practice of Luria-Delbruck fluctuation analysis: a review. Cancer Res 48:1060–1065

Kesten H (1989) Supercritical branching processes with countably many types and the size of random Cantor sets. In: Anderson TW, Athreya KB, Iglehart DL (eds) Probability, statistics, and mathematics: papers in honor of Samuel Karlin. Academic Press, Boston, pp 103–121

Kim JS, Kahng B, Kim D (2009) Disassortativity of random critical branching trees. Phy Rev E 79(6 Pt 2):067103

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998) Signatures of population expansion in microsatellite repeat data. Genetics 148:1921–1930

Kimmel M (1980) Integro-differential equations in cellular population dynamics. In: Kisielewicz M (ed) Functional differential systems and related topics. Mathematical Institute of the Polish Academy of Sciences, Zielona Gora, pp 149–160

Kimmel M (1980a) Cellular population dynamics. I: Model construction and reformulation. Math Biosci 48:211–224

Kimmel M (1980b) Cellular population dynamics. II: Investigation of solutions. Math Biosci 48:225–239

Kimmel M (1980c) Time discrete model of cell population dynamics. Syst Sci 6:343–363

Kimmel, M (1981) Branching processes in varying environment as dynamical systems with heredity. In: Kisielewicz M (ed) Functional differential systems and related topics II. Mathematical Institute of the Polish Academy of Sciences, Zielona Gora, pp 164–177

Kimmel M (1981) General theory of cell cycle dynamics based on branching processes in varying environment. In: Rotenberg M (ed) Biomathematics and cell kinetics. Elsevier, Amsterdam, pp 357–375

Kimmel M (1982) An equivalence result for integral equations with application to branching processes. Bull Math Biol 44:1–15

Kimmel M (1983) The point process approach to age- and time-dependent branching processes. Adv Appl Probab 15:1–20

Kimmel M (1985) Nonparametric analysis of stathmokinesis. Math Biosci 74:111–123

Kimmel M (1987) Metabolic events in the cell cycle of malignant and normal cells: a mathematical modeling approach. In: Thompson JR, Brown B (eds) Cancer modeling. Marcel Dekker, New York, pp 215–235

Kimmel, M (1994) Rapid genome evolution and cancer: a modeling perspective. Appl Math Comp Sci 4:163–177, Special Issue (Kimmel M, Swierniak A eds)

Kimmel M (1997) Quasistationarity in a branching model of division-within-division. In: Athreya KB, Jagers P (eds) Classical and modern branching processes. Volumes in mathematics and its applications, 84. Springer, New York, pp 157–164

Kimmel M, Arino O (1988) On active linear compartments. Math Comput Model 11:1189–1194

Kimmel M, Arino O (1991) Cell cycle kinetics with supramitotic control, two cell types and unequal division: A model of transformed embryonic cells. Math Biosci 105:47–79

Kimmel M, Arino O (1994) Two simple models of almost the same population with very different dynamics. Math Biosci 122:183–200

Kimmel M, Axelrod DE (1990) Mathematical models of gene amplification with applications to cellular drug resistance and tumorigenicity. Genetics 125:633–644

Kimmel M, Axelrod DE (1991) Unequal cell division, growth regulation and colony size of mammalian cells: A mathematical model and analysis of experimental data. J Theor Biol 153:157–180

Kimmel M, Stivers D (1994) A time-continuous branching process model of unstable gene amplification. Bull Math Biol 56:337–357

Kimmel M, Swierniak A (1982) On a certain optimal control problem related to the optimal chemotherapy of leukemia. Technical reports of the Silesian Technical University, Series Automation (Zeszyty Naukowe Politechniki Slaskiej, Seria Automatyka) 65:121–130. (in Polish, English abstract)

Kimmel M, Traganos F (1985) Kinetic analysis of drug induced G2 block in vitro. Cell Tissue Kinet 18:91–110
Kimmel M, Traganos F (1986) Estimation and prediction of cell cycle specific effects of anticancer drugs. Math Biosci 80:187–208
Kimmel M, Axelrod DE (1994) Fluctuation test for two-stage mutations: application to gene amplification. Mutat Res 306:45–60
Kimmel M, Arino O, Axelrod DE (1996) Backward/forward duality of branching processes and cell population dynamics. In: Martelli M, Cooke K, Cumberbatch E, Tang B, Thieme H (eds) Differential equations and applications to biology and to industry. World Scientific Publishing, Singapore, pp 233–240
Kimmel M, Axelrod DE (2002) Branching processes in biology. Springer, New York
Kimmel M, Axelrod DE, Wahl GM (1992) A branching process model of gene amplification following chromosome breakage. Mutat Res 276:225–239
Kimmel M, Darzynkiewicz Z, Staiano-Coico L (1986) Stathmokinetic analysis of human epidermal cells in vitro. Cell Tissue Kinet 19:289–304
Kimmel M, Darzynkiewicz Z, Arino O, Traganos F (1984) Analysis of a cell cycle model based on unequal division of metabolic constituents to daughter cells during cytokinesis. J Theor Biol 110:637–664
Kimmel M, Grossi A, Amuasi J, Vannucchi AM (1990) Non-parametric analysis of platelet lifespan. Cell Tissue Kinet 23:192–202
Kimmel M, Mathaes M (2010a) Griffiths-Pakes branching process as a model for evolution of Alu elements. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 179–189
Kimmel M, Mathaes M (2010b) Modeling neutral evolution of Alu elements using a branching process. BMC Genomics 11(Suppl 1):S11
Kimmel M, Traganos F, Darzynkiewicz Z (1983) Do all daughter cells enter the 'Indeterminate' ('A') state of the cell cycle? Analysis of stathmokinetic experiment on L1210 cells. Cytometry 4:192–201
Kirkwood TBL, Holliday R (1978) A stochastic model for the commitment of human cells to senescence. In: Valleron AJ, Macdonald PD (eds) Biomathematics and cell kinetics. Elsevier, Amsterdam, pp 161–172
Klebaner F (1988) Conditions for fixation of an allele in the density-dependent Wright-Fisher models. J Appl Probab 25:247–256
Klebaner F (1990) Conditions for the unlimited growth in multitype population size dependent Galton-Watson processes. Bull Math Biol 52:527–534
Klebaner F (1997) Population and density dependent branching process. In: Athreya KB, Jagers P (eds) Classical and modern branching processes. Springer, New York, pp 165–170
Klebaner F, Zeitouni O (1994) The exit problem for a class of density dependent branching systems. Ann Appl Probab 4:1188–1305
Klebaner FC (2010) Approximations in population-dependent branching processes. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin. pp 71–78
Klein B, Macdonald PDM (1980) The multitype continuous-time Markov branching process in a periodic environment. Math Sci 12:1–13
Knolle H (1988) Cell kinetic modelling and the chemotherapy of cancer. Springer-Verlag, Berlin
Komarova NL, Wodarz D (2007) Stochastic modeling of cellular colonies with quiescence: an application to drug resistance in cancer. Theor Popul Biol 72:523–538
Komarova NL, Cheng P (2006) Epithelial tissue architecture protects against cancer. Math Biosci 200:90–117
Komarova NL, Wodarz D (2013) Targeted cancer treatment in silico: small molecule inhibitors and oncolytic viruses. Birkhäuser, Boston
Koteeswaran P (1989) Estimating the age of a Galton-Watson process with binomial offspring distribution. Stoch Anal Appl 7:413–423

Kotenko JL, Miller JH, Robinson AI (1987) The role of asymmetric cell division in *Pteripodhyte* cell differentiation. I. Localized metal accumulation and differentiation in *Vittaria gemmae* and *Onoclea prothallia*. Protoplasma 136:81–95

Kowald A (1997) Possible mechanisms for the regulation of telomere length. J Mol Biol 273:814–825

Kozlov MV (1976) On the asymptotic probability of non-extinction for critical branching processes in random environment. Theor Probab Appl 21:791–804

Kozlov MV (1995) Conditional functional limit theorem for a critical branching processes in random environment. Dokl Ross Akad Nauk 344:12–15 (In Russian)

Krings M, Geisert H, Schmitz RW, Krainitzki H, Paabo S (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. Proc Natl Acad Sci U S A 96:5581–5585

Kuczek T (1984) Stochastic modeling for the bacterial cell cycle. Math Biosci 69:159–171

Kuczek T, Axelrod DE (1986) The importance of clonal heterogeneity and interexperiment variablity in modeling the eukaryotic cell cycle. Math Biosci 79:87–96

Kuczek T, Axelrod DE (1987) Tumor cell heterogeneity: divided-colony assay for measuring drug response. Proc Natl Acad Sci U S A 84:4490–4494

Kuczek T, Chan TCK (1988) Mathematical modeling for tumor resistance. J Natl Cancer I 80:146–147 Response: Goldie JH, Coldman AJ (1988) J Natl Cancer I 80:146–147

Kufe DW, Pollock RE, Weichselbaum RR, Bast Jr, et al (2003) Cancer Medicine, 6th ed. Decker, Inc. NY (Table of contents available online. Entire contents searchable online at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books&itool=toolbar)

Lai Y, Shinde D, Arnheim N, Sun F (2003) The mutation process of microsatellites during the polymerase chain reaction. J Comput Biol 10:143–155

Lai Y, Sun F (2004) Sampling distribution for microsatellites amplified by PCR: mean field approximation and its applications to genotyping. J Theor Biol 228:185–194

Lalam N (2006) Estimation of the reaction efficiency in polymerase chain reaction. J Theor Biol 242:947–953

Lalam N (2007) Statistical inference for quantitative polymerase chain reaction using a hidden Markov model: a Bayesian approach. Stat Appl Genet Mo B 6:1544–6115

Lalam N, Jacob C, Jagers P (2005) Estimation of the PCR efficiency based on a size-dependent modelling of the amplification process. Comptes Rendus de l'Académie des Sciences de Paris (Mathématique) 341:631–634

Lambert A (2006) Probability of fixation under weak selection: a branching process unifying approach. Theor Popul Biol 69:419–441

Lambert A (2008) Population dynamics and random genealogies. Stochastic Models 24(S1):45–163

Lambert A (2010) The contour of splitting trees is a Levy process. Ann Probab 38:348–395

Lambert A, Popovic L (2013) The coalescent point process of branching trees. Ann Appl Probab 23:99–144

Lapidus R (1984) Growth and division kinetics of asymmetrically dividing *Tetrahymena thermophilia*. J Theor Biol 106:135–140

Larson DD, Spangler EA, Blackburn EH (1987) Dynamics of teleomere length variation in *Tetrahymena thermophila*. Cell 50:477–483

Larsson S, Ryden T, Holst U, Oredsson S, Johansson M (2008a) Estimating the total rate of DNA replication using branching processes. Bull Math Biol 70:2177–2194

Larsson S, Ryden T, Holst U, Oredsson S, Johansson M (2008b) Estimating the variation in S phase duration from flow cytometric histograms. Math Biosci 213:40–49

Le Gall J-F (1999) Spatial branching processes, random snakes and partial differential equations. Birkhauser, Basel

Lea DE, Coulson CA (1949) The distribution of the numbers of mutants in bacterial populations. J Genet 49:264–265

Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A 109:E2274–2783

Leman SC, Levy F, Walker ES (2009) Modeling the spread of infectious disease using genetic information within a marked branching process. Stat Med 28:3626–3642

Levy MZ, Allsopp RC, Futcher AB, Greider CW, Harley CB (1992) Teleomere end-relplication problem and cell aging. J Mol Biol 225:951–960

Levy SB (1998) The challenge of antibiotic resistance. Sci Am 278(March):46–53

Lewontin RC (2000) The triple helix: gene, organism, and environment. Harvard University, Cambridge

Lieberman BS (2001) A test of whether rates of speciation were unusually high during the Cambrian radiation. P Roy Soc Lond B Bio 268:1707–1714

Lipow C (1975) A branching model with population size dependence. Adv Appl Probab 7:495–510

Loeffer M, Grosmann B (1991) A stochastic branching model with formation of subunits applied to the growth of intestinal crypts. J Theor Biol 150:175–191

Loeffler M, Roeder I (2002) Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models—a conceptual approach. Cells Tissues Organs 171:8–26

Loeffler M, Wichmann HE (1980) A comprehensive mathematical model of stem cell proliferation which reproduces most of the published experimental results. Cell Tissue Kinet 13:543–561

Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511

Ma WT, Sandri G. vH, Sarkar S (1992) Analysis of the Luria-Delbrück distribution using discrete convolution powers. J Appl Probab 29:255–267

Macdonald PDM (1978) Age distributions in the general cell kinetic model. In: Valleron AJ, Macdonald PDM (eds) Biomathematics and cell kinetics. Elsevier, Amsterdam, pp 3–20

Macken CA, Perelson AS (1985) Branching processes applied to cell surface aggregaton phenomena. Springer-Verlag, Berlin

Macken CA, Perelson AS (1988) Stem cell proliferation and differentiation. A multitype branching process model. Lecture notes in biomathematics, 76. Springer-Verlag, Berlin

Mackillop WJ (1986) Instrinsic versus acquired drug resistance. Cancer Treat Rep 70:817–818. Reply: Goldie JH, Coldman AJ (1986) Cancer Treat Rep 70:818

MacMillan HR, McConnell MJ (2011) Seeing beyond the average cell: branching process models of cell proliferation, differentiation, and death during mouse brain development. Theory Biosci 130:31–43

Maddox J (1992) Is molecular biology yet a science? Nature 355:201

Manna SS (2009) Branching process in a stochastic extremal model. Physical Review E 80(2 Pt 1): 021132

Marciniak-Czochra A, Kimmel M (2006) Dynamics of growth and signaling along linear and surface structures in very early tumors. Comput Math Methods Med 7:189–213

Marciniak-Czochra A, Stiehl T (2013) Mathematical models of hematopoietic reconstitution after stem cell transplantation. Contrib Math Comput Sci 4:191–206

Marshall JM (2008) A branching process for the early spread of a transposable element in a diploid population. J Math Biol 57:811–840

Mayhew MB, Robinson JW, Jung B, Haase SB, Hartemink AJ (2011) A generalized model for multi-marker analysis of cell cycle progression in synchrony experiments. Bioinformatics 27: i295–i303

Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson CB (2008) The molecular basis of cancer, 3rd edn. Saunders, Philadelphia

Mercier JF, Slater GW, Mayer P (2003) Solid phase DNA amplification: a simple Monte Carlo Lattice model. Biophys J 85:2075–2086

Metz JAJ, Diekmann O (1966) The dynamics of physiologically structured populations. Lecture notes in biomathematics, 68. Springer-Verlag, Berlin

Michor F, Nowak MA, Iwasa Y (2006) Evolution of resistance to cancer therapy. Curr Pharm Des 12:261–271
Mode CJ (1962) Some multi-dimensional birth and death processes and their applications in population genetics. Biometrics 18:543–567
Mode CJ (1971) Multitype branching processes. Elsevier, New York
Möhle M (2004) The time back to the most recent common ancestor in exchangeable population models. Adv Appl Probab 36:78–97
Möhle M (2008) Convergence to the coalescent and its relation to the time back to the most recent common ancestor. Fifth Colloquium on Mathematics and Computer Science. Discrete Mathematics and Theoretical Computer Science Proceedings Artificial Intelligence, pp 311–324, Nancy
Molina M (2010) Two-sex branching process literature. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 279–293
Morris VB, Taylor IW (1985) Estimation of nonproliferating cells in the neural retina of embryonic chicks by flow cytometry. Cytometry 6:375–380
Morris VB, Cowan R (1984) A growth curve of cell numbers in the neural retina of embryonic chicks. Cell Tissue Kinet 17:199–208
Morrow J (1970) Genetic analysis of azaguanine resistance in an established mouse cell line. Genetics 65:279–287
Moschopoulos P, Shpak M (2010) The distribution of family sizes under a time-homogeneous birth and death process. Commun Stat-Theor M 39:1761–1775
Moy SC (1967) Extensions of a limit theorem of Everett, Ulam and Harris on multitype branching processes to a branching process with countably many types. Ann Math Stat 38:992–999
Murnane JP, Yezzi MJ (1988) Association of high rate of recombination with amplification of dominant selectable gene in human cells. Somat Cell Mol Genet 14:273–286
Nagylaki T (1990) Models and approximations for random genetic drift. Theor Popul Biol 37:192–212
Nagylaki T (1992) Introduction to theoretical population genetics. Springer, Berlin
Nair KA, Mode CJ (1971) The reducible multidimensional age-dependent branching process. J Math Anal Appl 33:131–139
Navidi WC, Tavaré S, Arnheim N (2003) The roles of mutation rate and selective pressure on observed levels of the human mitochondrial DNA deletion $mtDNA^{4977}$. In: Goldstein DR (ed) Science and statistics: a festschrift for terry speed. Lecture notes monograph series 40. Institute for Mathematical Statistics, Beachwood, pp 247–258
Nedelman J, Rubinow SI (1980) Investigation into the experimental kinetic support of the two-state model of the cell cycle. Cell Biophys 2:207–231
Nedelman J, Downs H, Pharr P (1987) Inference for an age-dependent, multitype branching-process model of mast cells. J Math Biol 25:203–226 Erratum: J Math Biol 25:571
Neveu J (1975) Discrete-parameter martingales. Translated from the French by T. P. Speed. Revised edition. North-Holland Mathematical Library, Vol. 10. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York
Niccum BA, Poteau R, Hamman GE, Varada JC, Dshalalow JH, Sinden RR (2012) On an unbiased and consistent estimator for mutation rates. J Theor Biol 300:360–367
Nishiura H (2011) Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). Biomed Eng Online 10:15
Nordon RE, Ko KH, Odell R, Schroeder T (2011) Multi-type branching models to describe cell differentiation programs. J Theor Biol 277:7–18
Oates AC, Gorfinkiel N, González-Gaitán M, Heisenberg CP (2009) Quantitative approaches in developmental biology. Nat Rev Genet 10:517–530
O'Connell N (1993) Yule process approximation for the skeleton of a branching process. J Appl Probab 30:725–729

O'Connell N (1995) The genealogy of branching processes and the age of our most recent common ancestor. Ad Appl Probab 27:418–442
Olofsson P (1996) Branching processes with local dependencies. Ann Appl Probab 6:238–268
Olofsson P (2000) A branching process model of telomere shortening. Comm Statist Stochastic Models 16:167–177
Olofsson P, Bertuch AA (2010) Modeling growth and telomere dynamics in *Saccharomyces cerevisiae*. J Theor Biol 263:353–359
Olofsson P, Kimmel M (1999) Stochastic models of telomere shortening. Math Biosci 158:75–92
Olofsson P, Ma X (2011) Modeling and estimating bacterial lag phase. Math Biosci 234:127–131
Olofsson P, McDonald TO (2010) A stochastic model of cell cycle desynchronization. Math Biosci 223:97–104
Olofsson P, Shaw AC (2002a) Exact sampling formulas for multi-type Galton-Watson processes. J Math Biol 45:279–293
Olofsson P, Shaw AC (2002b) Exact sampling formulas for multi-type Galton-Watson processes. J Math Biol 45:279–293 Erratum 47:390 (2003)
Olofsson P, Schwalb O, Chakraborty R, Kimmel M (2001) An application of a general branching process in the study of the genetics of aging. J Theor Biol 213:547–57
Olovnikov AM (1973) A theory of marginotomy. J Theor Biol 41:181–190
Orlando DA, Iversen Jr ES, Hartemink AJ, Haase SB (2009) A branching process model for flow cytometry and budding index measurements in cell synchrony experiments. Ann Appl Stat 3:1521–1541
Pakes AG (1992) Limit theorems for the numbers of rare mutants: a branching process model. Ad Appl Prob 24:778–794
Pakes AG (1993) Explosive Markov branching processes: entrance laws and limiting behaviour. Ad Appl Prob 25:737–756
Pakes AG (1994) On the recognition & structure of probability generating functions. Research Report, Department of Mathematics, The University of Western Australia. Nedlands, WA, Australia, pp 1–29
Pakes AG (2000) Biological applications of branching processes. Research Report, Department of Mathematics and Statistics, The University of Western Australia. Nedlands, WA, Australia
Pakes AG (2003) Biological applications of branching processes. In: Shanbhag DN, Rao CR (eds) Handbook of statistics 21: stochastic processes: modelling and simulation. North-Holland, Amsterdam, pp 693–773
Pakes AG, Dekking FM (1991) On family trees and subtrees of simple branching processes. J Theor Probab 4:353–369
Pakes AG, Trajstman AC (1985) Some properties of continuous-state branching processes, with applications to Bartoszynski's virus model. Ad Appl Prob 17:23–41
Pelengaris S, Khan M (eds) (2013) The Molecular Basis of Cancer, 2nd ed. Wiley-Blackwell, Hoboken
Peterson JA (1984) Analysis of variability in albumin content of sister hepatoma cells and model for geometric phenotypic variability (Quantitative Shift Model). Somat Cell Mol Genet 10:345–357
Pharr PN, Nedelman J, Downs HP, Ogawa M, Gross AJ (1985) A stochastic model for mast cell proliferation in culture. J Cell Physiol 125:379–386
Piau D (2001) Processus de branchement en champ moyen et réaction PCR. [Mean-field branching process and the PCR reaction]. Ad Appl Prob 33:391–403
Piau D (2002) Mutation-replication statistics of polymerase chain reactions. J Comput Biol 9:831–847
Poil SS, van Ooyen A, Linkenkaer-Hansen K (2008) Avalanche dynamics of human brain oscillations: relation to critical branching processes and temporal correlations. Hum Brain Mapp 29:770–777
Polanski A, Kimmel M, Swierniak A (1997) Qualitative analysis of the infinite model of drug resistance evolution. In: Arino O, Axelrod D, Kimmel M (eds) Advances in mathematical population dynamics - molecules, cells and man. World Scientific, Singapore, pp 595–612

Polanski A, Swierniak A, Duda Z (1993) Multiple solutions to the TPBVP arising in optimal scheduling of cancer chemotherapy. Conference Proceedings 1993, IEEE International Conference on Systems, Man and Cybernetics 4:5–8
Pollak E (1974) Survival probabilities and extinction times for some multitype branching processes. Adv Appl Probab 6:446–462
Pollak E (1996) On the calculation of the effective population size by a method based on the theory of branching processes. J Indian Soc Agr Stat 49:47–66, iii
Pollak E (2000) The effective population size of some age-structured populations. Math Biosci 168:39–56
Puck TT, Steffen J (1963) Life cycle analysis of mammalian cells. Part I. Biophys J 3:379–397
Ramo P, Kesseli J, Yli-Harja O (2006) Perturbation avalanches and criticality in gene regulatory networks. J Theor Biol 242:164–70
Rass L (2007) Convergence results for contact branching processes. Math Biosci 205:59–76
Richards RI, Sutherland GR (1994) Simple repeat DNA is not replicated simply. Nat Genet 6:114–116
Ridout MS, Cole DJ, Morgan BJ, Byrne LJ, Tuite MF (2006) New approximations to the Malthusian parameter. Biometrics 62:1216–1223
Rigney DR (1981) Correlation between the ages of sibling cell cycle events and a test of the "transition probabability" cell cycle model. In: Rotenberg M (ed) Biomathematics and cell kinetics. Elsevier, Amsterdam, pp 157–166
Rincon A, Alonso JA, Sanz L (2009) Reduction of supercritical multiregional stochastic models with fast migration. Acta Biotheor 57:479–500
Rittgen W (1983) Controlled branching processes and their applications to normal and malignant haematopoiesis. Bull Math Biol 45:617–626
Roeder I, Loeffler M (2002) A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. Exp Hematol 30:853–861
Roeder I, Lorenz R (2006) Asymmetry of stem cell fate and potential impact of the niche. Stem Cell Rev 2:171–180
Roeder I, Hergberg M, Horn M (2009) An "Age"-structured model of hematopoietic stem cell organization with application to chronic myeloid leukemia. Bull Math Biol 71:602–626
Rosen R (1986) Role of mathematical modeling in protocol formulation in cancer chemotherapy. Cancer Treat Rep 40:1461–1462. Reply: Coldman AJ, Goldie JH (1986) Cancer Treat Rep 70:1461–1462
Roze D, Barton NH (2006) The Hill-Robertson effect and the evolution of recombination. Genetics 173:1793–811
Sagitov S (1988) Limit behavior of reduced critical branching processes. (Russian) Doklady Akademii Nauk SSSR 303, 47–49. Translation in Soviet Mathematics Doklady 38:(1989) 488–491
Sagitov S (2011) Unpublished slides of Serik Sagitov, presented at a workshop at Luminy on April 26, 2011
Sagitov S (2013) Linear-fractional branching processes with countably many types. Stoch Processes Appl 123:2940–2956
Saiki RK, Gelfand DH, Stoffel S, Scharf SJ et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491
Sawyer S (1976) Branching diffusion processes in population genetics. Ad Appl Prob 8:659–689
Schinazi RB (2006) The probability of treatment induced drug resistance. Acta Biotheor 54:13–19
Sehl M, Zhou H, Sinsheimer JS, Lange KL (2011) Extinction models for cancer stem cell therapy. Math Biosci 234:132–46
Seneta E, Tavaré S (1983) Some stochastic models for plasmid copy number. Theor Popul Biol 23:241–256
Seneta E (2006) Non-negative matrices and Markov chains. Springer Series in Statistics No. 21, Springer, New York

Sennerstam R (1988) Partition of protein (mass) to sister cell pairs at mitosis: a re-evaluation. J Cell Sci 90:301–306
Sennerstam R, Strömberg J-O (1984) A comparative study of the cell cycles of nullipotent and mulitpotent embryonal carcinoma cell lines during exponential growth. Dev Biol 103:221–229
Sennerstam R, Strömberg J-O (1988) Evidence for an intraclonal random shift between two types of cell cycle times in an embryonal carcinoma cell line. J Theor Biol 131:151–162
Sennerstam R, Strömberg J-O (1995) Cell cycle progression: Computer simulation of uncoupled subcycles of DNA replication and cell growth. J Theor Biol 175:177–189
Sennerstam R, Strömberg J-O (1996) Exponential growth, random transitions and progress through the G1 phase: Computer simulation of experimental data. Cell Prolif 29:609–622
Serra MC, Haccou P (2007) Dynamics of escape mutants. Theor Popul Biol 72:167–178
Shaw CA (2000) Genealogical methods for multitype branching processes with applications in biology. Ph.D dissertation (Kimmel, M., advisor). Department of Statistics, Rice University, Houston, TX
Shiri T, Welte A (2011) Modelling the impact of acute infection dynamics on the accumulation of HIV-1 mutations. J Theor Biol 279:44–54
Silvestre DA, Fontanari JF (2007) Preservation of information in a prebiotic package model. Phys Rev E 75(5 Pt 1):051909
Skipper HE (1983) The forty-year old mutation theory of Luria and Delbrück and its pertinence to cancer chemotherapy. Adv Cancer Res 40:331–363
Smith JA, Martin L (1973) Do cells cycle? Proc Natl Acad Sci U S A 70:1263–1267
Spătaru A (1989) Properties of branching processes with denumerably many types. Revue Roumaine de Mathématiques Pures et Appliquées. (Romanian Journal of Pure and Applied Mathematics) 34:747–759
Staiano-Coico L, Hajjar DP, Hefton JM, Hajjar K, Kimmel M (1988) Interaction of arterial Cells: III. Stathmokinetic analyses of smooth muscle cells cocultured with endothelial cells. J Cell Physiol 134:485–490
Staudte RG (1992) A bifurcating autoregression model for cell lineage data with varying generation means. J Theor Biol 156:183–195
Staudte RG, Guiguet M, d'Hooghe MC (1984) Additive models for dependent cell populations. J Theor Biol 109:127–146
Staudte RG, Huggins RM, Zhang J, Axelrod DE, Kimmel M (1997) Estimating clonal heterogeneity and interexperiment variability with the bifurcating autoregression model for cell lineage data. Math Biosci 143:103–121
Stewart FM, Gordon DM, Levin BR (1990) Fluctuation analysis: the probability distribution of the number of mutants under different conditions. Genetics 124:175–185
Stiehl T, Ho AD, Marciniak-Czochra A (2014) The impact of CD34+ cell dose on engraftment after stem cell transplantations: personalized estimates based on mathematical modeling. Bone Marrow Transplant 49:30–37
Stigler SM (1970) Estimating the age of a Galton-Watson branching process. Biometrika 57:505–512
Stivers DN, Kimmel M (1996a) A continuous-time, multi-type generational inheritance branching process model of cell proliferation with clonal memory. Nonlinear World 3:385–399
Stivers DN, Kimmel M (1996b) On the clonal inheritance model of cell proliferation. In: Proceedings of the First World Congress of Nonlinear Analysts, Tampa, Florida, August 1992. Walter de Gruyter, Berlin, pp 3401–3408
Stivers DN, Kimmel M, Axelrod DE (1996) A discrete-time, multi-type generational inheritance branching process model of cell proliferation. Math Biosci 137:25–50
Stoneking M, Sherry ST, Redd AJ, Vigilant L (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. Philos T Roy Soc B 337:167–175
Strome ED, Wu X, Kimmel M, Plon SE (2008) Heterozygous screen in *Saccharomyces cerevisiae* identifies dosage-sensitive genes that affect chromosome stability. Genetics 178:1193–1207

Subramanian VG, Duffy KR, Turner ML, Hodgkin PD (2008) Determining the expected variability of immune responses using the cyton model. J Math Biol 56:861–892
Swierniak A, Kimmel M (1984) Optimal control application to leukemia chemotherapy protocols design. Technical reports of the Silesian Technical University, Series Automation (Zeszyty Naukowe Politechniki Slaskiej, Seria Automatyka) 73:261–277. (In Polish, English abstract)
Swierniak A, Kimmel M (1991) Cancer cell synchronization and recruitment as optimal control problems. In: Proceedings of 13th World IMACS Congress. Dublin 3:1461–1462
Swierniak A, Polanski A, Kimmel M (1996) Control problems arising in chemotherapy under evolving drug resistance. In: Preprints of the 13th World Congress of International Federation of Automatic Control, Volume B, pp 411–416
Tachikawa M (2008) Fluctuation induces evolutionary branching in a mathematical model of ecosystems. PLoS One 3(12):e3925
Taïb Z (1987) Labelled Branching Processes with Applications to Neutral Evolution Theory. Ph.D. thesis, Chalmers University of Technology, Sweden
Taïb Z (1992) Branching processes and neutral evolution. Lecture notes in biomathematics, 93. Springer-Verlag, Berlin
Taïb Z (1993) A note on modeling the dynamics of budding yeast populations using branching process. J Math Biol 31:805–815
Taïb Z (1995) Branching processes and functional-differential equations determining steady-size distributions in cell populations. J Appl Prob 32:1–10
Tan WY (1982) On the distribution theories for the number of mutants in cell populations. SIAM J Appl Math 42:719–730
Tan WY (1983) On the distribution of the number of mutants at the hypoxanthine-quanine phosphoribosyl transferase locus in Chinese hamster ovary cells. Math Biosci 67:175–192
Tan WY, Yan XW (2010) A new stochastic and state space model of human colon cancer incorporating multiple pathways. Biol Direct 5:26–42
Taneyhill DE, Dunn AM, Hatcher MJ (1999) The Galton-Watson branching process as a quantitative tool in parasitology. Parasitol Today 15:159–165
Tannock I (1978) Cell kinetics and chemotherapy: a critical review. Cancer Treat Rep 62:1117–1133
Tavaré S (1980) Time-reversal and age distribution. 1. Discrete Markov case. J Appl Probab 17:33–46
Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetics models. Theor Popul Biol 26:119–164
Therneau TM, Solberg LA Jr, Jenkins RB (1989) Modeling megakaryocyte development as a branching process. Comput Math Appl 18:959–964
Till JE, McCulloch EA, Siminovitch L (1964) A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. Proc Natl Acad Sci U S A 51:29–36
Tltsy T, Margolin BH, Lum K (1989) Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbrück fluctuation analysis. Proc Natl Acad Sci U S A 86:9441–9445
Tomasetti C, Levy D (2010) Role of symmetric and asymmetric division of stem cells in developing drug resistance. Proc Natl Acad Sci U S A 107:16766–16771
Tomasetti C, Vogelstein B, Parmigiani G (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. Proc Natl Acad Sci U S A 110:1999–2004
Traganos F, Kimmel M (1990) The stathmokinetic experiment: a single-parameter and multiparameter flow cytometric analysis. In: Darzynkiewicz Z, Crissman HA (eds) Methods in cell biology, Vol. 23. Academic Press, New York, pp 249–270
Tyrcha J (1988) Asymptotic stability in a generalized probabilistic/deterministic model of the cell cycle. J Math Biol 26:465–475
Tyson JJ (1987) Size control of cell division. J Theor Biol 126:381–391
Tyson JJ, Hannsgen KB (1985) Global asymptotic stability of the size distribution in probabilistic models of the cell cycle. J Math Biol 22:61–68

Tyson JJ, Hannsgen KB (1985) The distributions of cell size and generation time in a model of the cell cycle incorporating size control and random transitions. J Theor Biol 113:29–62
Tyson JJ, Hannsgen KB (1986) Cell growth and division: a deterministic/probabilistic model of the cell cycle. J Math Biol 23:231–246
Tyson J, Garcia-Herdugo G, Sachsenmaier W (1979) Control of nuclear division in *Physarum polycephalum*. Exp Cell Res 119:87–98
Varshaver NB, Marshak MI, Shapiro NI (1983) The mutational origin of serum independence in Chinese hamster cells in vitro. Int J Cancer 31:471–475
Vazquez A (2006) Polynomial growth in branching processes with diverging reproductive number. Phys Rev Lett 96(3):038702
Venter JC (2001) The sequence of the human genome. Science 291:1304–1351
Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson A (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A 86:9350–9354
Vigilant L, Stoneking R, Harpending H, Hawkes K, Wilson A (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
Vogel H, Niewisch H, Matioli G (1969) Stochastic development of stem cells. J Theor Biol 22:249–270
Waghmare SK, Bruschi CV (2005) Differential chromosome control of ploidy in the yeast Saccharomyces cerevisie. Yeast 22:625–639
Wang D, Zhao C, Cheng R, Sun F (2000) Estimation of the mutation rate during error-prone polymerase chain reaction. J Comput Biol 7:143–158
Waugh, W.A.O'N (1981) Application of the Galton-Watson process to the kin number problem. Advances in Applied Probability 13:631–649
Webb GF (1987) Random transitions, size control, and inheritance in cell population dynamics. Math Biosci 85:71–91
Webb GF (1989) Alpha- and beta-curves, sister-sister and mother-daughter correlations in cell population dynamics. Comput Math Appl 18:973–984
Weber GF (2007) Molecular mechanisms of cancer. Springer, New York
Weinberg RA (2014) The biology of cancer, 2nd ed. Garland Science, New York
Wellard C, Markham J, Hawkins ED, Hodgkin PD (2010) The effect of correlations on the population dynamics of lymphocytes. J Theor Biol 264:443–449
Weiss G, von Haeseler A (1997) A coalescent approach to the polymerase chain reaction. Nucleic Acids Res 25:3082–3087
Wilke CO (2003) Probability of fixation of an advantageous mutant in a viral quasispecies. Genetics 163:467–474
Wilson AC, Cann RL (1992) Recent African genesis of humans. Sci Am 266(4):68–73
Windle B, Draper BW, Yin YX, O'Gorman S, Wahl GM (1991) A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. Genes Dev 5:160–714
Wodarz D, Komarova NL (2005) Computational biology of cancer: lecture notes and mathematical modeling. World Scientific Publishers, Singapore
Wu X, Strome ED, Meng Q, Hastings PJ, Plon SE, Kimmel M (2009) A robust estimator of mutation rates. Mutat Res 661:101–109
Xiong X, Boyett JM, Webster RG, Stech J (2009) A stochastic model for estimation of mutation rates in multiple-replication proliferation processes. J Math Biol 59:175–191
Xu K, Verghese G, Doerschuk P (2005) Evaluating estimates of Markov models of sequence evolution through simulation. Conf Proc IEEE Eng Med (Baltim) Biol 1:808–812
Yakovlev A, Yanev NM (2006) Branching stochastic processes with immigration in analysis of renewing cell populations. Math Biosci 203:37–63
Yanev NM (2010) Branching processes in cell proliferation kinetics. In: González M et al. (eds) Workshop on branching processes and their applications. Lecture notes in statistics—proceedings 197. Springer-Verlag, Berlin, pp 159–178

Yanev NM, Yakovlev AY (1985) On the distribution of marks over a proliferating cell population obeying the Bellman-Harris branching process. Math Biosci 75:159–173
Yates A, Chan C, Strid J, Moon S et al (2007) Reconstruction of cell population dynamics using CFSE. BMC Bioinformatics 8:196
Yule UG (1924) A mathematical theory of evolution based on conclusions of Dr. J.C. Willis, F.R.S. Philosophical Transactions of the Royal Society of London. Series B 213:21–87
Zakian VA (1995) Telomeres: Beginning to understand the end. Science 270:1601–1607
Zakian VA (1996) Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. Annu Rev Genet 30:141–172
Zang Y, Garré M, Gjuracic K, Bruschi CV (2002) Chromosome V loss due to centromere knock-out or MAD2-deletion is immediately followed by restitution of homozygous diploidy in Saccharomyces cerevisiae. Yeast 19:553–564
Zheng Q (1999) Progress of a half century in the study of the Luria-Delbrück distribution. Math Biosci 162:1–32
Zheng Q (2002) Statistical and algorithmic methods for fluctuation analysis with salvador as an implementation. Math Biosci 176:237–252
Zheng Q (2005) Update on estimation of mutation rates using data from fluctuation experiments. Genetics 171:861–864
Zilman A, Ganusov VV, Perelson AS (2010) Stochastic models of lymphocyte proliferation and death. PLoS ONE 5(9):e12775
Zubkov AM (1975) Limiting distribution for the distance to the nearest common ancestor. Theor Probab Appl 20:602–612

Author Index

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0

Subject Index

M. Kimmel, D. E. Axelrod, *Branching Processes in Biology,*
Interdisciplinary Applied Mathematics 19, DOI 10.1007/978-1-4939-1559-0